Psychology, Public Policy, and Law
2005, Vol. 11, No. 1, 3–41
Copyright 2005 by the American Psychological Association
1076-8971/05/$12.00 DOI: 10.1037/1076-8971.11.1.3

CRITERIA-BASED CONTENT ANALYSIS
A Qualitative Review of the First 37 Studies

Aldert Vrij
University of Portsmouth

Statement Validity Assessment (SVA) is used to assess the veracity of child witnesses' testimony in trials for sexual offences. The author reviewed the available SVA research. Issues addressed include the accuracy of Criteria-Based Content Analysis (CBCA; part of SVA), interrater agreement between CBCA coders, frequency of occurrence of CBCA criteria in statements, the correlations between CBCA scores and (i) interviewer's style and (ii) interviewee's age and social and verbal skills, and issues regarding the Validity Checklist (another part of SVA). Implications for the use of SVA assessments in criminal courts are discussed. It is argued that SVA evaluations are not accurate enough to be admitted as expert scientific evidence in criminal courts but might be useful in police investigations.

To date, Statement Validity Assessment (SVA) is probably the most popular instrument for assessing the veracity of child witnesses' testimony in trials for sexual offences (Vrij, 2000). SVA assessments are accepted as evidence in some American courts (Ruby & Brigham, 1997) and in criminal courts in several West European countries, such as Sweden (Gumpert & Lindblad, 1999), Germany (Köhnken, 2002), and the Netherlands (Lamers-Winkelman & Buffing, 1996). SVA should be used more widely according to Honts (1994), who argued that its validity has been conclusively demonstrated, and Raskin and Esplin (1991a, 1991b) and Zaparniuk, Yuille, and Taylor (1995) have pressed for the use of the SVA procedure in North American criminal courts. Others, however, are more skeptical (Brigham, 1999; Davies, 2001; Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997; Rassin, 1999; Ruby & Brigham, 1997; Wells & Loftus, 1991).

Statement analysis, initially less systematic than the current SVA procedure, has been applied by German experts in criminal court cases since the 1950s (Steller & Boychuk, 1992), but research into the accuracy of statement analysis originated much later. It has been pointed out that it was after more than 30 years of forensic practice in German courts that the first German study to validate statement analysis was published (Steller, 1989). Obviously, more empirical research regarding the validity of the procedure is needed (Doris, 1994). Research papers testing the accuracy of statement analysis appeared in English for the first time in the late 1980s, and to my knowledge, 37 studies have been published and/or presented (in English) at conferences to date. The findings of these studies are discussed in this article. At the core of SVA is Criteria-Based Content Analysis (CBCA; Berliner & Conte, 1993), and therefore, perhaps unsurprisingly, most of these 37 studies have focused on the accuracy of CBCA analyses. Previous reviews of CBCA literature have been published (Horowitz, 1991;

Correspondence concerning this article should be addressed to Aldert Vrij, Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth PO1 2DY, United Kingdom. E-mail: aldert.vrij@port.ac.uk

Lamb, Sternberg, Esplin, Hershkowitz, & Orbach, 1997; Pezdek & Taylor, 2000; Ruby & Brigham, 1997; Tully, 1999; Vrij, 2000). However, the present is the most comprehensive review as it includes more studies than previous reviews and addresses a greater number of issues, such as interrater agreement rates; the frequency of occurrence of the individual CBCA criteria in statements; the effect of age, verbal ability, social ability, and interview style on CBCA scores; and several aspects related to the Validity Checklist (another part of SVA).

Statement Validity Assessment: History

Statement analysis originated in Germany and Sweden. It is perhaps not surprising that a technique has been developed to verify whether or not a child has been sexually abused. It is often difficult to determine the facts of a sexual abuse case because often there is no medical or physical evidence. Frequently, the alleged victim and the defendant give contradictory testimony, and often, there are no independent witnesses to give an objective version of events. Thus, the perceived credibility of the defendant and of the alleged victim is important. The alleged victim is in a disadvantageous position if he or she is a child as adults have a tendency to mistrust statements made by children (Ceci & Bruck, 1995).

The German psychologists Arntzen (1982) and Undeutsch (1982) and the Swedish psychologist Trankell (1972) suggested various criteria that they believed could be used to assess the veracity of statements. The first to describe such a list of criteria was Undeutsch (1967). The hypothesis underlying these criteria is that "truthful, reality-based accounts differ significantly and noticeably from unfounded, falsified, or distorted stories" (Undeutsch, 1982, p. 44). Undeutsch emphasized that apart from examining these criteria, other aspects need to be taken into consideration as well to form a final opinion about the veracity of a statement, such as the degree to which the statement is consistent with information from other sources (Gumpert & Lindblad, 1999). With the help of others, Gunter Köhnken and Max Steller took statement analysis a step further, refining Undeutsch's criteria and integrating them into a formal assessment procedure they called SVA (Köhnken & Steller, 1988; Raskin & Esplin, 1991b; Raskin & Steller, 1989; Raskin & Yuille, 1989; Steller, 1989; Steller & Boychuk, 1992; Steller & Köhnken, 1989; Yuille, 1988b). SVA consists of three elements: a semistructured interview, CBCA, and an evaluation of the CBCA outcomes.

Stage 1: The Semistructured Interview

The first stage of SVA is a semistructured interview in which the child provides his or her own account of the allegation. A key element is that the child tells his or her own story without any influence from the interviewer. Several researchers have designed special interview techniques based on psychological principles to obtain as much information as possible from children in a free narrative style (Bull, 1992, 1995, 1998; Davies, Westcott, & Horan, 2000; Hershkowitz, 1999, 2001; Lamb, Sternberg, & Esplin, 1994, 1998; Lamb, Sternberg, Orbach, Hershkowitz, & Esplin, 1999; Memon & Bull, 1999; Milne & Bull, 1999; Raskin & Esplin, 1991b; Sternberg, Lamb, Esplin, Orbach, & Hershkowitz, 2002) without inappropriate prompts or suggestions. Appropriate prompts (such

as "What happened next?") or questions (e.g., "You just mentioned a man. What did he look like?") are part of such techniques.

Stage 2: Criteria-Based Content Analysis

The interviews are audiotaped and transcribed, and the transcripts are used for the second part of SVA: CBCA. Trained evaluators judge the presence or absence of 19 criteria (see Figure 1). CBCA is based on the hypothesis, originally stated by Undeutsch (1967), that a statement derived from memory of an actual experience differs in content and quality from a statement based on invention or fantasy, known as the *Undeutsch hypothesis* (Steller, 1989). The presence of each criterion strengthens the hypothesis that the account is based on genuine personal experience. In other words, truthful statements have more of the elements measured by CBCA than do false statements. A theoretical foundation for the Undeutsch hypothesis was presented by Köhnken (1989, 1996, 1999), who proposed that both cognitive and motivational factors influence CBCA scores.

With regard to cognitive factors, it is assumed that the presence of several

General Characteristics

1. Logical structure
2. Unstructured production
3. Quantity of details

Specific Contents

4. Contextual embedding
5. Descriptions of interactions
6. Reproduction of conversation
7. Unexpected complications during the incident
8. Unusual details
9. Superfluous details
10. Accurately reported details misunderstood
11. Related external associations
12. Accounts of subjective mental state
13. Attribution of perpetrator's mental state

Motivation-Related Contents

14. Spontaneous corrections
15. Admitting lack of memory
16. Raising doubts about one's own testimony
17. Self-deprecation
18. Pardoning the perpetrator

Offense-Specific Elements

19. Details characteristic of the offense

Figure 1. Content criteria for statement analysis. From "Criteria-Based Content Analysis" (p. 221), by M. Steller and G. Köhnken, in *Psychological Methods in Criminal Investigation and Evidence*, ed. by D. C. Raskin, 1989, New York: Springer-Verlag. Copyright 1989 by Springer Publishing Company, Inc., New York 10036. Used by permission.

criteria (Criteria 1–13; see Figure 1) are likely to indicate genuine experiences as they are typically too difficult to fabricate. Therefore, statements that are coherent and consistent (*logical structure*), whereby the information is not provided in a chronological time sequence (*unstructured production*), and that contain a significant amount of detail (*quantity of detail*) are more likely to be true. Regarding details, accounts are more likely to be truthful if they include *contextual embeddings* (references to time and space: "He approached me for the first time in the garden during the summer holidays"), *descriptions of interactions* ("The moment my mother came into the room, he stopped smiling"), *reproduction of speech* (speech in its original form: "And then he asked, 'Is that your coat?'"), *unexpected complications* (elements incorporated in the statement that are somewhat unexpected, e.g., the child mentions that the perpetrator had difficulty with starting the engine of his car), *unusual details* (details that are uncommon but meaningful, e.g., a witness who describes that the man she met had a stutter), and *superfluous details* (descriptions that are not essential to the allegation, e.g., a witness who describes that the perpetrator was allergic to cats). Another criterion that might indicate truthfulness is when a witness speaks of details that are beyond the horizon of his or her comprehension, for example, when he or she describes the adult's sexual behavior but attributes it to a sneeze or to pain (*accurately reported details misunderstood*). Finally, possible indicators of truthfulness are if the child reports details that are not part of the allegation but are related to it (*related external associations,* e.g., a witness who describes that the perpetrator talked about the women he had slept with and the differences between them), describes his or her feelings or thoughts experienced at the time of the incident (*accounts of subjective mental state*), or describes the perpetrator's feelings, thoughts, or motives during the incident (*attribution of perpetrator's mental state:* "He was nervous, his hands were shaking").

Other criteria (Criteria 14–18; see Figure 1) are more likely to occur in truthful statements for motivational reasons. Truthful persons are not as concerned with impression management as deceivers. Compared with truth tellers, deceivers are more keen to try to construct a report that they believe will make a credible impression on others, and so they leave out information that, in their view, will damage their image of being a sincere person (Köhnken, 1999). As a result, a truthful statement is more likely to contain information that is inconsistent with the stereotypes of truthfulness. The CBCA list includes five of these so-called contrary-to-truthfulness-stereotype criteria (Ruby & Brigham, 1998): *spontaneous corrections* (corrections made without prompting from the interviewer ("He wore black trousers, no, sorry, they were green"), *admitting lack of memory* (expressing concern that some parts of the statement might be incorrect: "I think," "maybe," "I am not sure," etc.), *raising doubts about one's own testimony* (anticipated objections against the veracity of one's own testimony: "I know this all sounds really odd"), *self-deprecation* (mentioning personally unfavorable, self-incriminating details: "Obviously, it was stupid of me to leave my door wide open because my wallet was clearly visible on my desk"), and *pardoning the perpetrator* (making excuses for the perpetrator or failing to blame him or her, such as a girl who says she now feels sympathy for the defendant who possibly faces imprisonment).

The final criterion relates to *details characteristic of the offense*. This criterion

is present if a description of events is typical for the type of crime under investigation (e.g., a witness describes feelings that professionals know are typical for victims of, say, incestuous relationships).

Stage 3: Evaluation of the CBCA Outcome

CBCA scores might be affected by factors other than the veracity of the statement. Take, for example, the age of the interviewee. Cognitive abilities and command of language develop throughout childhood, making it gradually easier to give detailed accounts of what has been witnessed (Davies, 1991, 1994a; Fivush, Haden, & Adam, 1995). Therefore, all sorts of details are less likely to occur in the statements of young children. Also, children under 8 years old may have difficulty in viewing the world from somebody else's perspective (Flavell, Botkin, Fry, Wright, & Jarvis, 1968); thus, Criterion 13 (accounts of perpetrator's mental state) is unlikely to occur in the statements of young children. Finally, younger children have less developed metacognitive and metamemorial capabilities (i.e., knowing whether or not they know or remember an answer; Walker & Warren, 1995), so they are less likely to be aware of gaps in their memories (Criterion 15).

A Validity Checklist has been developed consisting of issues that are thought to be relevant and so worth examining as they might affect CBCA scores. Detailed descriptions of the issues mentioned in the Validity Checklist have been provided by Raskin and Esplin (1991b), Steller (1989), Steller and Boychuk (1992), and Yuille (1988b). Slightly different versions of the Validity Checklist exist (different authors have used somewhat different versions). The Validity Checklist presented below is the one published by Steller and colleagues (Steller, 1989, Table II, used here with kind permission of Springer Science and Business media; see also Steller & Boychuk, 1992). SVA evaluators consider the following issues: (a) appropriateness of language and knowledge (mental capability of the child); (b) appropriateness of affect shown by the interviewee; (c) interviewee's susceptibility to suggestion; (d) evidence of suggestive, leading, or coercive questioning; (e) overall adequacy of the interview; (f) motives to report, for example, whether the interviewee's relationship with the accused or with other people involved suggests possible motives for a false allegation; (g) context of the original disclosure or report, for example, whether there are questionable elements in the context of the original disclosure; (h) pressures to report falsely, such as indications that others suggested, coached, pressured, or coerced the interviewee to make a false report; (i) consistency with the law of nature, that is, whether the described events are unrealistic; (j) consistency with other statements, that is, whether there are major elements of the statement that are inconsistent or contradicted by another statement made by this interviewee; and (k) consistency with other evidence, for example, whether there are major elements in the statement that are contradicted by reliable physical evidence or other concrete evidence. Henceforth, I refer to such issues as *external factors.* In the third stage of the SVA procedure, evaluation of the CBCA outcome, the evaluator systematically addresses each of the external factors mentioned in the checklist and explores and considers alternative interpretations of the CBCA outcomes.

Three of these factors have been addressed in CBCA research: (a) age of the

interviewee, (b) interviewer's style, and (c) coaching of the interviewee. These are discussed in this review. A fourth external factor, verbal skills of the interviewee, has been examined in CBCA research but is not included in the Validity Checklist. This factor is also discussed.

The Literature Review

Studies Included in the Review

All published articles and book chapters that appeared in a literature search (PsycLIT, using the search terms *Criteria-Based Content Analysis, Statement Validity Assessment*, and *Statement Validity Analysis*) were included in this review. In addition, I included all known CBCA/SVA conference papers, including those that have been reported previously in the SVA literature (e.g., in Bradford, 1994; Bybee & Mowbray, 1993; Davies, 2001; Horowitz, 1991; Köhnken, Schimossek, Aschermann, & Höfer, 1995; Steller, 1989; Vrij, 2000). Finally, Boychuk's (1991) unpublished field study has been included. This study has received extensive coverage in the SVA literature (e.g., in Horowitz, 1991; Lamb, Sternberg, Esplin Hershkowitz, & Orbach, 1997; Lamers-Winkelman, 1995; and Ruby & Brigham, 1997). The studies included in the review are indicated with an asterisk in the reference list.

Type of Studies

In an attempt to validate the assumptions of CBCA, two types of studies have been conducted. In field studies, statements made by persons in actual cases of alleged sexual abuse have been examined, whereas in experimental laboratory studies, statements of participants who lied or told the truth for the sake of the experiment have been assessed. Each paradigm has its advantages, and the one's strength is the other's weakness. The statements assessed in field studies have clear forensic relevance as these are statements derived from real-life cases. However, it is often difficult to establish the truth or falsity of these statements beyond doubt.

Typically, criteria such as confession, polygraph results, and conviction have been used to establish whether a statement is actually true or false. The problem is that these criteria are often not independent from the quality of the statement and, therefore, from CBCA scores. For example, statements were classified as doubtful if the judge dismissed the charges in studies conducted by Esplin, Boychuk, and Raskin (1988) and Boychuk (1991). However, a dismissal might simply be the result of the child being unable to express convincingly to the judge or jury what he or she had experienced; it does not necessarily imply that the child is lying.

Another criterion often used to establish whether a statement is actually true or false is a confession (Craig, Scheibe, Raskin, Kircher, & Dodd, 1999). However, if the only evidence against the guilty defendant is the incriminating statement of the child, which is often the case in sexual abuse cases, it is unlikely that the perpetrator will confess to the crime if the incriminating statement is of poor quality because the perpetrator's main motivation for confessing to a crime is the perception that the evidence against him or her is strong (Moston, Stephen-

son, & Williamson, 1992). On the other hand, if a false incriminating statement is persuasive and judged to be truthful by a CBCA expert, the chances of the innocent defendant's obtaining an acquittal decrease dramatically, and if there is no chance of avoiding a guilty verdict, it may be beneficial to plead guilty to obtain a reduced penalty (Steller & Köhnken, 1989). In summary, poor-quality (e.g., unconvincing) statements decrease the likelihood of obtaining a confession, and high-quality (e.g., convincing) statements increase the likelihood of obtaining a confession, regardless of whether a statement is truthful or fabricated.

Good field studies establish whether the statement is actually true or false on the basis of criteria that are independent from the witness statement, such as DNA evidence and medical evidence. However, that type of evidence is often not available in real-life cases in which CBCA assessments are conducted (Steller & Köhnken, 1989). For a discussion about difficulties in establishing whether a statement is true or false in studies of sexual abuse, see Horowitz et al. (1996).

In experimental laboratory studies, there is no difficulty establishing whether a statement is actually true or false, but experimental situations typically differ from real-life situations. Recalling a film someone has just seen (a paradigm sometimes used in laboratory studies) is different from describing a sexual abuse experience. Therefore, because of this lack of ecological validity, Undeutsch (1984) believed that laboratory studies are of little use in testing the accuracy of SVA analyses. Clearly, researchers should attempt to make laboratory studies as realistic as possible and should try to create situations that mimic elements of actual child sexual abuse cases.

Steller (1989) has argued that experiences of sexual abuse are characterized by three important elements: (a) personal involvement, (b) negative emotional tone of the event, and (c) extensive loss of control over the situation. The first element could be easily introduced into an experimental study; the latter two elements are more difficult because of ethical constraints. A popular paradigm in experimental CBCA research therefore is to invite participants to give an account of a negative event that they have experienced, such as giving a blood sample, being bitten by a dog, and so on, or to give a fictitious account of such an event that they have not actually experienced. Obviously, the experimenter needs to establish whether the story is actually true or fictitious, for example, by checking with the participants' parents, although this does not always happen in experimental research (see, e.g., Ruby & Brigham, 1998).

Different studies have used different paradigms, and the paradigms used are listed in Table 1. A distinction is made between field studies and laboratory studies. In the laboratory studies, a further distinction is made between studies in which respondents actually participated in an event and were asked to tell the truth or lie about that event afterwards (active), studies in which they were shown a video and then asked to tell the truth or lie about that video (video), studies in which they watched a staged event and then were asked to tell the truth or lie about that event (staged), and studies in which they were asked to tell a truthful or fictitious story about a previous negative experience in their life (memory).

As mentioned before, CBCA was developed to evaluate statements from children who are witnesses or alleged victims in sexual abuse cases. Many authors still describe CBCA as a technique developed solely to evaluate statements made by children in sexual offense trials (see, e.g., Honts, 1994; Horowitz et al. 1997).

Table 1
Differences Between Truth Tellers and Liars on CBCA Criteria

Authors	Age (years)	Event	Status	CBCA criterion 1	2	3	4	5
Field studies								
Boychuk (1991)	4–16	Field	Victim	>	>	>	>	>
Craig et al. (1999)	3–16	Field	Victim					
Esplin et al. (1988)	3–15	Field	Victim	>	>	>	>	>
Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav (1997)	4–13	Field	Victim	—	>	>	>	>
Parker & Brown (2000)	Adult	Field	Victim	—	>	>	—	>
Laboratory studies								
Akehurst et al. (2001)	7–11/ adult	Active	Na	>	—	>	—	>
Colwell et al. (2002)	Adult	Staged	Witness	>				
Höfer et al. (1996)	Adult	Active	Na	>	—	>	>	—
Köhnken et al. (1995)	Adult	Video	Witness	—	>	>		—
Landry & Brigham (1992)	Adult	Memory	Victim	<		>	>	>
Porter & Yuille (1996)	Adult	Active	Suspect	>		>		
Porter et al. (1999)	Adult	Memory	Victim			—		
Ruby & Brigham (1998)	Adult	Memory	Victim	<	>	—	<	>
Santtila et al. (2000)	7–14	Memory	Victim	—	>	>	—	—
Sporer (1997)	Adult	Memory	Victim	>	—	—	>	—
Steller et al. (1988)	6–11	Memory	Victim	>		>	>	—
Tye et al. (1999)	6–10	Active	Witness	—	>	>	>	—
Vrij, Edward, et al. (2000)	Adult	Video	Witness	—	—	>	>	—
Vrij & Heaven (1999)	Adult	Video	Witness			—		
Vrij, Kneller, & Mann (2000)[a]	Adult	Video	Witness	—	—	>	—	—
Vrij et al. (in press)	5–15/ adult	Active	Witness/ suspect	>		>	>	>
Winkel & Vrij (1995)	8–9	Video	Witness	>	>	>	>	>
Total (support/total number of studies ratio)				10/19	9/14	16/20	11/16	9/17
Total support in percentages				53	64	80	69	53

Note. CBCA = Criteria-Based Content Analysis; Na = participants participated in an activity but were neither victims nor suspects; > = verbal characteristic occurs more frequently in truthful than in deceptive statements; < = verbal characteristic occurs more frequently in deceptive than in truthful statements; — = no relationship between the verbal characteristic and lying/truth telling. Blank cells indicate that the verbal characteristic was not investigated.
[a]Uninformed liars only.

CBCA criterion														
6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
>	>	>	>	—	>	>	—	>	—	—	—	>	—	
														>
>	>	>	>	—	>	>	>	>	>	—	—	>	>	>
>	—	—	—	—	—	—	—	—						>
>	—	—	—	—	—	>	>	—	—	—	—	—		
>	—	—	—			>	—	—						>
>	>	—	—			>	—	—						>
—	—	—	—	—			—	—	>	—				
>	—	>	>			>	<	>	>	>	—			>
—	—	—	—		—	—		—	>					
									>					
—	>	>	>		<	<		>	>	—	<			<
>	—	>	—	—	—	—	>	—						
—	—	—	—	—	—	—	—							
—	>	>	>	>	>	—	—	—	—		<	—		
>	—	—	>			—	—	—						>
>		>	—			—	>	>	—	—		—		>
										>				
	—	>	—				>	>	—	—				>
>						—		—	—	—				>
		>	—		>			—	—	—				>
11/16	5/15	9/17	6/17	1/8	4/10	6/15	5/14	6/17	6/13	2/11	0/6	2/5	1/2	11/12
69	33	53	35	12	40	40	36	35	46	18	0	40	50	92

Others, however, have advocated the additional use of the technique to evaluate the testimonies of adults who talk about issues other than sexual abuse (Köhnken et al., 1995; Porter & Yuille, 1996; Ruby & Brigham, 1997; Steller & Köhnken, 1989). These authors have pointed out that the underlying Undeutsch hypothesis is restricted neither to children, witnesses, and victims nor to sexual abuse. To shed light on this issue, I have also indicated in Table 1 whether the statements were derived from children or adults and whether they were victims, witnesses, or suspects. Participants who discussed a negative life event they had experienced have been labelled as victims.

Differences Between Truth Tellers and Liars in CBCA Scores: Field Studies

Several researchers have conducted field studies without examining differences in CBCA scores between truthful and false accounts (Anson, Golding, & Gully, 1993; Buck, Warren, Betman, & Brigham, 2002; Davies et al., 2000; Hershkowitz, Lamb, Sternberg, & Esplin, 1997; Lamers-Winkelman & Buffing, 1996). These researchers have examined the impact of factors such as age or interview style on CBCA scores, and their findings are discussed in the Validity Checklist section. Others have examined the impact of veracity on CBCA scores.

The first CBCA field study ever presented was Esplin et al.'s (1988) study. A trained CBCA evaluator rated the statements. If a criterion was not present in the statement, it received a score of 0; if it was present, it received a score of 1; and if it was strongly present, it received a score of 2. Hence, total CBCA scores could range between 0 and 38. The results were striking. The confirmed cases received a mean CBCA score of 24.8, and the doubtful statements received a mean score of 3.6. Moreover, the distributions of scores of the confirmed and doubtful groups did not show a single overlap. The highest score in the doubtful group was 10 (one child received that score, and three children obtained a score of 0), whereas the lowest score in the confirmed group was 16 (one child obtained that score, and the highest score was 34). When differences between the two groups on each criterion were assessed, differences between the doubtful and confirmed groups emerged for 16 out of 19 criteria, all in the expected direction. That is, the criteria were more often present in the confirmed cases than in the doubtful cases, which strongly supports the Undeutsch hypothesis (see Table 1). However, Esplin et al.'s study has been heavily criticized (Wells & Loftus, 1991).

The problems with Esplin et al.'s (1988) study include the facts that only one evaluator scored the transcripts, that the effect could simply have been an age effect, and that the decision about what really had happened was not based on independent case facts. Wells and Loftus (1991) pointed out that the differences between the two groups could have been caused by age differences between these groups. Indeed, the children in the confirmed group were older (9.1 years) than the children in the doubtful group (6.9 years). Moreover, the doubtful group included eight statements from children who were younger than 5 years old, whereas the confirmed group contained only one statement from a child under 5 years old. Second, independent criteria used for the doubtful cases in the study were *judicial dismissal, no prosecution, no confessions made by the defendant,* and *persistent*

denial by the accused. As identified earlier, none of these criteria are independent case facts.

In her subsequent study, Boychuk (1991) addressed some of these criticisms. Statements of 75 children between the ages of 4 and 16 years old were analyzed by three raters who were masked with regard to case disposition. She also included in her sample, apart from confirmed and doubtful groups, a third group: likely abused. The likely abused were those without medical evidence but with confessions by the accused or criminal sanctions from a superior court. Unfortunately, in all of her analyses, including the one presented in Table 1, she combined the confirmed group and the likely abused group. By assessing differences between the two remaining groups on each criterion, Boychuk found fewer significant differences than Esplin and colleagues (1988) had (see Table 1), but all 13 differences found were in the expected direction. That is, the criteria were more often present in the confirmed cases than in the doubtful cases, which again supports the Undeutsch hypothesis.

CBCA assessments were carried out to assess the veracity of adult rape allegations in a field study published by Parker and Brown (2000). Differences were found on several criteria, and all differences were in the expected direction (see Table 1). However, this study also had serious methodological problems. For example, the criteria for establishing the actual veracity of the statements, *convincing evidence of rape* (when no information was given as to what was meant by this) and *corroboration in the legal sense and with either a suspect being identified or charged,* are either too vague or not independent case facts. Also, only one evaluator examined most of the cases, and it is unclear whether that person was masked with regard to the facts of the case or if she or he had any background information about the cases she or he was asked to assess.

In a better controlled field study, Lamb, Sternberg, Esplin, Hershkowitz, Orbach, and Hovav (1997) selected and analyzed the statements of 98 alleged victims of child sexual abuse (aged 4–12 years) and included only cases in which there was (a) evidence of actual physical contact between a known accused and the child and (b) an element of corroboration present. Using these selection criteria meant that many other cases needed to be disregarded as the initial sample consisted of 1,187 interviews.[1] They found fewer significant differences than Boychuk (1991) and Esplin et al. (1988) partly because not all 19 criteria were included in the assessment. However, again, all differences were in the expected direction, that is, the criteria were more often present in the plausible group than in the implausible group. Like Esplin and colleagues, Lamb et al. also calculated the mean CBCA scores of their two groups. If a criterion was not present in the statement, it received a score of 0; if it was present, it received a score of 1. Only 14 criteria were used in this study, which meant that the total CBCA score could vary between 0 and 14. Significantly more criteria were present in the confirmed

[1]This is a problem field researchers typically face when they use stringent selection criteria. For example, Anson et al. (1993) could use only 23 cases that fit their selection criteria out of a sample of 466 cases. An important issue is whether a small, selective sample means that the sample is unrepresentative. The fact that the sample is small might affect generalization, but using stringent selection criteria should not affect representativeness as there are no good reasons to believe that strong independent corroborative evidence would change the nature of a child's disclosure.

cases (6.74) than in the doubtful cases (4.85). This difference, however, is much smaller than the difference found by Esplin and colleagues.

Craig et al. (1999) examined 48 statements from children between the ages of 3 and 16 years old who were alleged victims of sexual abuse. A statement was classified as confirmed if the accused made a confession and/or failed a polygraph test. A statement was classified as highly doubtful if the child provided a detailed and credible recantation and/or the accused passed a polygraph test, that is, when the polygraph test suggested that the accused was innocent. In other words, this study also did not establish independent case facts. The average CBCA score of the confirmed cases (7.2) was slightly higher than the average score of the doubtful cases (5.7). Only 14 criteria were used, and the scores could vary between 0 and 14. Only total CBCA scores were examined.[2]

Differences Between Truth Tellers and Liars in CBCA Scores: Laboratory Studies

Compared with most field studies, the laboratory studies revealed fewer differences between liars and truth tellers per study (see Table 1). Almost all differences, however, were in the expected direction, with the criteria occurring more frequently in truthful reports than in deceptive reports, supporting the Undeutsch hypothesis. This consistent support for CBCA criteria is striking when compared with research into nonverbal indicators of deception, in which the findings are much more erratic (see Vrij, 2000, for a review of such research). Almost all of the findings that deviated from the general pattern were obtained in Landry and Brigham's (1992) and Ruby and Brigham's (1998) studies. Several explanations for this might be possible. They used raters who were trained for only 45 min in CBCA scoring, and it is doubtful whether people could be considered CBCA trained after such a short training period. Also, judges were exposed to very short statements, on average 255 words, whereas the CBCA method has been developed for use on longer statements (Raskin & Esplin, 1991b). Finally, in Landry and Brigham's study, some judges did not read the transcripts of the statements (common CBCA procedure) but watched videotaped statements instead. CBCA experts are typically not in favor of assessing videotaped statements as watching a videotape might distract the CBCA assessor from his or her assessment task (Köhnken, 1999). People also perceive the frequency of occurrence of verbal criteria differently in different presentation modes: They

[2]In a case study published by Orbach and Lamb (1999), the accuracy of a statement provided by a 13-year-old sexual abuse victim could be established with more certainty and in greater detail than in most other studies. Information given by the victim during the interview was compared with an audiotaped record of that incident. The child had told her mother that her grandfather had sexually molested her on several occasions, but the mother did not believe the allegations. When one day the grandfather entered the bathroom while the child was listening to music played on an audiotape recorder, she pressed the record button and recorded the sexually abusive incident that was about to unfold. Orbach and Lamb conducted a CBCA analysis on the statement and found that 10 out of 14 criteria they assessed were present in the statement. Obviously, the results of a study in which only one (truthful) statement is examined do not say much about the validity of CBCA. Also, the fact that the child knew that there was audiotaped evidence of the incident might have influenced her statement in an unspecified manner. Nevertheless, the nature and strength of the corroborative evidence make the study worth mentioning.

typically believe that such criteria are more present when watching a video than when reading a text (Strömwall & Granhag, 2003). Moreover, research has demonstrated that people are better at detecting truths and lies when they read a transcript than when they watch a video (see DePaulo, Stone, & Lassiter, 1985, for a review). In other words, these studies deviated considerably from the normal CBCA procedure on certain points that may have affected the CBCA judgments.

Furthermore, Table 1 shows that in both children's and adults' narratives, the criteria emerged more frequently in truthful reports. Age differences were tested for directly by including statements from both adults and children in experiments by Akehurst, Köhnken, and Höfer (2001) and Vrij, Akehurst, Soukara, and Bull (2002). They both found higher total CBCA scores for truth tellers than for liars in both age groups (children vs. adults); however, they did not examine age differences on the separate criteria, and the results presented in Table 1 are the combined scores for adults and children.

Some criteria occurred more frequently in statements from innocent suspects than in statements from guilty suspects in a study by Porter and Yuille (1996). Vrij et al. (2002) are the only researchers to have directly compared statements of suspects and witnesses. They found a higher total CBCA score for truth tellers than for liars in both suspects and witnesses but did not examine differences on each criterion.

These findings support the assumption that CBCA ratings are not restricted to statements of victims and children about sexual abuse but could be used in different contexts and with different types of interviewee. However, one should keep in mind that CBCA assessments can be used only for statements that have been provided in interviews in which free recall was stimulated and prompting was kept at a minimum. Such an interview style rarely occurs in police interviews with suspects, which means that conducting CBCA assessments on suspects' statements would probably often be inappropriate.

Finally, the expected differences were found in CBCA scores between liars and truth tellers in all experimental research paradigms—actual involvement, watching a video, statements derived from memory, and so on—which is a further indication that differences in CBCA scores are rather robust.

A look at the empirical support for each of the 19 criteria shows that Criterion 3 (quantity of detail) received the most support. The amount of detail was calculated in 20 studies, and in 16 of those studies (80%), truth tellers included significantly more details in their accounts than liars (see the bottom of Table 1). Unstructured production (Criterion 2), contextual embeddings (Criterion 4), and reproduction of conversation (Criterion 6) all received strong support as well. The so-called motivational criteria, Criteria 14 to 18, received less support than most cognitive criteria (1–13). In fact, Criterion 17, self-deprecation, has received no support at all to date. This criterion has been examined in six studies. In two studies, a significant difference between liars and truth tellers appeared, and both times, the criterion appeared less often in the truthful statements. Berliner and Conte (1993) pointed out that Criteria 14 to 16 require the witness to exhibit a lack of confidence in the account as evidence for truthfulness. This, they noted, suggests by implication that confidence diminishes the likelihood of truthfulness, which is an implication they find disputable. As can be seen in Table 1, several researchers did not examine Criteria 15 to 19 either because of interrater reliabil-

ity concerns (Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997) or because they believed these criteria are theoretically unrelated to the basic memory concept embodied in the Undeutsch hypothesis (Raskin & Esplin, 1991b). Accurately reported details misunderstood (Criterion 10) and raising doubts about one's own testimony (Criterion 16) received little support too, perhaps, as is shown below, because these criteria are not frequently present in statements.

The hypothesis that truth tellers would obtain a higher total CBCA score than liars was examined in 12 studies. In 11 out of these 12 studies (92%), the hypothesis was supported.

Interrater Reliability Scores

Several authors have reported interrater reliability scores using different methods: (a) proportion agreement rates, (b) correlations (Pearson or Spearman), (c) Cohen's kappas, or (d) Maxwell's random error coefficient of agreement (RE). Percentage agreement can be inflated by chance (Maxwell, 1977), so it is preferable to use chance-corrected statistics. Cohen's kappa is such a statistic. However, kappa is known to be inaccurate when the base rates significantly diverge from .50 (Spitznagel & Helzer, 1985), that is, when a criterion is hardly present in any of the statements or when a criterion is present in almost all statements. It has been argued that Maxwell's RE is the best statistic to use in those circumstances (Maxwell, 1977). Several researchers scored the presence or absence of criteria on Likert-type scales, and in such cases, calculating agreement rates is inappropriate. In such cases, correlations were used instead.

Regardless of the method used, a score of .50 or higher could be considered as adequate reliability (Anson et al., 1993; Fleiss, 1981). According to Fleiss (1981), scores between .60 and .75 could be considered as good and scores over .75 as excellent. Table 2 shows the interrater agreement scores in CBCA studies and the percentage of studies in which a good interrater agreement rate (.60 or higher) was obtained. For most criteria, good interrater agreements were obtained in the majority of studies (exceptions are Criterion 2 [unstructured production] and Criterion 14 [spontaneous corrections]). Many interrater agreement rates were above .75, and interestingly, all three studies in which interrater agreement was calculated for the total CBCA score fell in this excellent range. Only the interrater agreement for the total CBCA score fell into this excellent category in Vrij, Akehurst, Soukara, and Bull's (in press) study. These findings suggest that total CBCA scores are more reliable than scores for the individual criteria.

Frequency of Occurrence of CBCA Criteria

Several researchers have examined how often the criteria were present in statements. Table 3 shows a review of their findings. Although frequency of occurrence scores have been calculated in both field studies and laboratory studies, the findings of field studies are probably more relevant as the occurrence depends on an event that someone has witnessed. For example, if participants in a laboratory study have witnessed a video in which no unusual details occurred, the frequency of occurrence of unusual details in those witnesses' statements is likely to be very low. Therefore, I discuss only the field studies here (but see Table 3 for percentages found in laboratory studies).

Table 2
Interrater Agreement Scores

Authors	Age (years)	Event	Status	CBCA criterion 1	2	3	4
Field studies							
Anson et al. (1993)	4–12	Field	Victim	.65	.13	.65	.48
Boychuk (1991)	4–16	Field	Victim	> .83	> .83	> .83	> .83
Buck et al. (2002)	2–14	Field	Victim	.67	.79	.32	.77
Craig et al. (1999)	3–16	Field	Victim	> .72	> .72	> .72	> .72
Horowitz et al. (1997)[a]	2–19	Field	Victim	.77	.50	.58	.75
Laboratory studies							
Akehurst et al. (2001)	7–11/ adult	Active	Na	.34	.35	.68	.42
Colwell et al. (2002)	Adult	Staged	Witness	.83			
Höfer et al. (1996)	Adult	Active	Na				
Porter & Yuille (1996)	Adult	Active	Suspect	> .80		> .80	
Porter et al. (1999)	Adult	Memory	Victim			> .70	.24
Santtila et al. (2000)	7–14	Memory	Victim	> .63	> .63	Nc	> .63
Vrij, Edward, et al. (2000)	Adult	Video	Witness	.55	.65	.90	.85
Vrij, Kneller; & Mann (2000)[b]	Adult	Video	Witness	> .87	.53	> .87	> .87
Vrij et al. (2004)	5–15/ adult	Active	Witness/ suspect	.49	.08	.56	.76
Vrij et al. (2001a)	Adult	Video	Witness	1.00	.51	.90	.88
Winkel & Vrij (1995)	8–9	Video	Witness	> .73	> .73	> .73	> .73
Total (good[c] interrater scores/ number of studies ratio)			11/14	6/12	10/13	10/13	9/12
Total (percentage of good interrater agreement scores)			79	50	77	77	75

Note. CBCA = Criteria-Based Content Analysis; Na = participants participated in activity but were neither victims nor suspects; Nc = interrater agreement was not calculated; MAX = Maxwell's random error coefficient of agreement; KAPPA = Cohen's kappa; COR = Pearson or Spearman correlations; AGREE = proportion agreement. Blank cells indicate that the verbal characteristic was not investigated.
[a]First occasion scores only. [b]Uninformed liars only. [c]*Good* was defined as .60 or higher.

(*table continues*)

As can be seen in Table 3, the frequency of occurrence of criteria differs widely for each criterion. In particular, Criterion 1 (logical structure), Criterion 3 (quantity of details), Criterion 4 (contextual embeddings), and Criterion 19 (details characteristic of the offense) are often present, whereas Criterion 10 (accurately reported details misunderstood), Criterion 16 (raising doubts about

Table 2 (*continued*)

Authors	CBCA criterion							
	5	6	7	8	9	10	11	12
Field studies								
Anson et al. (1993)	.13	.65	.56	.39	.48	.83	.22	.13
Boychuk (1991)	> .83	> .83	> .83	> .83	> .83	> .83	> .83	> .83
Buck et al. (2002)	.69	.52	.79	.73	.59	.69	.55	.65
Craig et al. (1999)	> .72	> .72	> .72	> .72	> .72	> .72	> .72	> .72
Horowitz et al. (1997)[a]	.65	.71	.57	.48	.37	.83	.52	.57
Laboratory studies								
Akehurst et al. (2001)	.44	.67	.49	.33	.55	−.04		.62
Colwell et al. (2002)								
Höfer et al. (1996)								
Porter & Yuille (1996)		> .80	> .80	> .80		> .80		> .80
Porter et al. (1999)								
Santtila et al. (2000)	> .63	.87	> .63	> .63	> .63	> .63	> .63	> .63
Vrij, Edward, et al. (2000)	.90	.97		.77	.69			.58
Vrij, Kneller; & Mann (2000)[b]	> .87		> .87	> .87	> .87			
Vrij et al. (2004)	.55	.52	.30	.05				.68
Vrij et al. (2001a)	.82	.79		.52				
Winkel & Vrij (1995)	> .73			> .73	> .73		> .73	
Total (good[c] interrater scores/ number of studies ratio)	9/11	6/10	8/13	7/11	6/7	5/8	7/10	8/11
Total (percentage of good interrater agreement scores)	82	60	62	64	86	63	70	73

memory), and Criterion 17 (self-deprecation) rarely occur in statements (typically in less than 10% of the statements). The latter three criteria are also those with the least support for the Undeutsch hypothesis (see Table 1). Several researchers have examined age differences in the frequency of occurrence of CBCA criteria. These results are discussed later.

CBCA criterion								
13	14	15	16	17	18	19	Total	Type
.83	.39	.22	1.00	.74	1.00	.22		MAX
> .83	> .83	> .83	> .83	> .83	> .83	> .83		KAPPA
.90	.46	.53	.94	.88	.86	.71		MAX
> .72	> .72							KAPPA
.67	.24	.39	.96	.88	.89	.75	> .78	MAX
.58	.35	.02	−.06				Nc	COR
								AGREE
							.78	COR
.67		> .80	> .80					COR
		> .70						COR
> .63	> .63							COR
.71	.54	.89	.70		1.00		Nc	COR
> .87	> .87	> .87	> .87				Nc	AGREE
.20	.57	.66	.68				.85	COR
.25	.51	.50	.14					KAPPA
	> .73	> .73	> .73				Nc	KAPPA
6/13	7/12	8/10	4/4	5/5	3/4	3/3		
46	58	80	100	100	75	100		

Correct Classifications of Truth Tellers and Liars on the Basis of Their CBCA Scores

Are trained evaluators better than laypersons? The first issue for discussion is whether there is any evidence that classifications of truth tellers and liars based on CBCA scores are better than classifications made by laypersons. Several

Table 3
Frequency of Occurrence of the CBCA Criteria (in Percentages)

Authors	Age (years)	Event	Status	CBCA criterion 1	2	3	4	5
Field Studies								
Anson et al. (1993)	4–12	Field	Victim	91	70	74	74	48
Boychuk (1991) confirmed	4–16	Field	Victim	100	100	100	96	66
Boychuk (1991) doubtful	4–16	Field	Victim	68	40	48	44	12
Buck et al. (2002)	2–14	Field	Victim	77	13	79	97	30
Esplin et al. (1988) true	3–15	Field	Victim	100	95	100	100	100
Esplin et al. (1988) false	3–15	Field	Victim	55	15	55	35	30
Horowitz et al. (1997)[a]	2–19	Field	Victim	87	71	77	89	32
Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav (1997) plausible	4–13	Field	Victim	100	76	97	82	62
Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav (1997) implausible	4–13	Field	Victim	100	46	77	46	23
Lamers-Winkelman & Buffing (1996)	2–11	Field	Victim	82	45	100	39	31
Total (percentage of occurrence in field studies)[b]				86	55	85	75	41
Laboratory studies								
Landry & Brigham (1992)	Adult	Memory	Victim	86		84	66	71
Tye et al. (1999) true	6–10	Active	Witness	92	92	83	75	42
Tye et al. (1999) false	6–10	Active	Witness	63	44	13	6	13
Vrij & Heaven (1999) true	Adult	Video	Witness					
Vrij & Heaven (1999) false	Adult	Video	Witness					
Vrij et al. (2001a)	Adult	Video	Witness	100	44	Nc	33	7

Note. CBCA = Criteria-Based Content Analysis; Nc = interrater agreement was not calculated. Blank cells indicate that the verbal characteristic was not investigated.
[a]First occasion scores only. [b]The total scores were calculated as follows. Criteria 1–14 were scored in a total of 543 statements, and the percentages presented are the percentage of occurrence in these 543 statements (e.g., Criterion 1 was present in 468 [86%] out of 543 statements). Criteria 15–19 were assessed in a total of 445 statements, and the percentages presented are the percentage of occurrence in these 445 statements (e.g., Criterion 15 was present in 203 [46%] out of 445 statements).

studies in which CBCA experts or laypersons judged children's statements are discussed in Vrij (2002a), culminating in the tentative conclusion that CBCA experts were better than laypersons. However, the studies included in that review used either CBCA experts or laypersons as judges, so a direct comparison could not be made. Also, there was probably a confound in those studies. In CBCA studies, experts made their judgments on the basis of the written transcripts,

CBCA criterion													
6	7	8	9	10	11	12	13	14	15	16	17	18	19
61	24	19	57	9	28	61	17	20	37	0	15	9	65
74	64	52	50	12	42	64	10	86	54	14	16	36	76
20	8	8	24	0	0	24	4	36	52	8	4	12	56
46	14	9	28	8	25	30	5	46	31	1	5	2	43
70	70	95	100	5	90	90	40	100	75	10	25	55	100
0	0	0	5	5	0	30	0	10	35	0	0	5	30
51	29	22	36	10	56	40	18	42	49	1	5	94	97
74	33	41	4	8	4	49	16	26					
46	23	15	0	15	8	38	23	8					
33	16	23	22	3	42	79	15	21	50	4	3	67	68
50	27	26	29	7	32	51	13	40	46	4	7	45	69
30	43	58	54			85	39	13	4	9	13		
67	0	0	75			25	17	8					
13	0	0	31			13	6	0					
										27			
										5			
20		27					99	80	5	3			

whereas in studies with laypersons, judgments were typically made on the basis of watching videotapes with interviewees. As mentioned before, people are better at detecting truths and lies when they read a transcript than when they watch a video (DePaulo et al., 1985).

Several researchers have examined the impact of CBCA training directly by including trained and untrained judges in their samples. Unfortunately, little is

known about what kind of training is actually required to become a CBCA expert. According to Raskin and Esplin (1991b), a 2- or 3-day workshop is advisable, whereas Köhnken (1999) recommended a 3-week training course. Moreover, nobody has tested whether such training actually works.[3] Although it is unclear how much training is required, it sounds reasonable to suggest that it should be a rather extensive training program. Making CBCA/SVA assessments is never a straightforward task. During CBCA coding, 19 criteria, some of which are difficult to score, need to be taken into consideration. After the CBCA coding, the impact of numerous external factors on the final statement needs to be assessed carefully (Steller, 1989; Wegener, 1989). It is impossible to do all this appropriately without extensive training, and even a 2- or 3-day workshop might be too short.

All studies that have examined the impact of CBCA training on accuracy scores clearly fall short of this 2- or 3-day-workshop requirement (Akehurst, Bull, & Vrij, 1998; Köhnken, 1987; Landry & Brigham, 1992; Ruby & Brigham, 1998; Santtila, Roppola, Runtti, & Niemi, 2000; Steller, Wellershaus, & Wolf, 1988; Tye, Amato, Honts, Kevitt, & Peters, 1999). The shortest training session (45 min) was given by Landry and Brigham (1992) and Ruby and Brigham (1998), though at 90 min, Steller et al.'s (1988) training session did not last much longer. Akehurst et al.'s (1998) session lasted 2 hr, whereas Köhnken (1987) and Santtila et al. (2000) did not provide information about the length of their training sessions. However, their sessions might well have been of similar length because the content of the training sessions used in those two studies strongly resembled the training sessions used in the other studies mentioned so far. In a typical study, trainees are given a handout with information about CBCA criteria. A trainer then explains the criteria in more depth and provides some examples. Trainees are then asked to rate one or a few exercise statements, and their ratings are discussed. The training session was slightly different in Tye et al.'s (1999) study as, rather than training judges specifically for their experiment, they used a panel of people who were previously trained in CBCA (no information was given about the training these previously trained judges had received).

The results of these training studies are mixed. Several researchers have found that trained judges were better at distinguishing between truths and lies than lay evaluators (Landry & Brigham, 1992; Steller et al., 1988; Tye et al., 1999). Some found no training effect (Ruby & Brigham, 1998; Santtila et al., 2000), and others found that training made judges worse at distinguishing between truths and lies (Akehurst et al., 1998; Köhnken, 1987). It is probably not fair to discredit CBCA training on the basis of these findings given the lack of depth of these training

[3]In the only field study related to this issue, Gumpert, Lindblad, and Grann (2002a) compared expert testimony reports prepared by professionals who had a statement analysis background with reports prepared by a more clinically oriented group often employed within child and adolescent psychiatry. They found that the reports of the statement analysis group were generally of higher quality (see Gumpert, Lindblad, & Grann, 2002b, for how quality was measured). Unfortunately, this study does not reveal anything about the effectiveness of CBCA training. As the authors acknowledged, they did not assess the accuracy of the recommendations made in the reports. Moreover, the groups could have differed in other respects besides training.

sessions. All one can conclude is that providing judges with such short training programs has an unpredictable effect on the ability to detect truths and lies.

Different ways of calculating accuracy rates. In CBCA research, accuracy rates—the correct classifications of liars and truth tellers—are computed in three different ways. First, CBCA scores might be subjected to statistical analyses, typically, discriminant analysis. Although this is a sound way of calculating accuracy rates, CBCA experts do not use such analyses in real life.

A second method is by asking CBCA experts to make truth–lie classifications. This method is more realistic as this is what happens in real life. However, it is also highly subjective because a classification depends on an assessor's own interpretation of a statement. The obvious problem with subjectivity is generalization. There is no guarantee that two different CBCA experts who judge the same statements will make the same decisions. In other words, the accuracy rate obtained by one expert in a CBCA study does not predict the accuracy rate obtained by a second expert in the same study.

A third method is by using decision rules. In this case, the truth–lie judgment is based on fixed rules, such as "the first five criteria should be present plus two others" (Zaparniuk et al., 1995). The advantage of this method is that it is objective: Different assessors who apply the same decision rule will obtain the same accuracy rates. However, it has serious shortcomings. As I mentioned earlier, CBCA scores depend on factors other than veracity, such as age and interview style, and these factors are ignored when such decision rules are used. CBCA experts are therefore opposed to the use of decision rules (Steller & Köhnken, 1989), but researchers nevertheless sometimes use them, even in field studies (Parker & Brown, 2000).

Accuracy rates in field and laboratory studies. The only field study in which accuracy rates were reported and a very high 90% overall accuracy rate was found was conducted by Parker and Brown (2000; see also Table 4). Not a single overlap between CBCA scores of confirmed and unconfirmed cases was found by Esplin et al. (1988). All scores for the unconfirmed cases were lower than any of the scores for the confirmed cases, which implies that Esplin et al. found an even higher (100%) accuracy rate. Although both studies showed tremendous support for the accuracy of CBCA assessments, as discussed earlier, both studies also had methodological flaws. I therefore prefer to disregard these results.

Regarding the remaining studies in which accuracy rates were reported (all laboratory studies), overall accuracy rates in those studies varied from 65% to 90%, with the exception of Landry and Brigham (1992), who obtained a lower accuracy rate. I have already given several reasons to explain their exceptional findings—short training, short statements, watching videotapes. In addition to this, the judges were advised to use a decision rule in which more than five criteria present equaled a good indication of high credibility, which is not what CBCA experts typically do. If one disregards their findings, Table 4 reveals that accuracy rates for truths varied between 53% and 89% and accuracy rates for lies between 60% and 100%. The average accuracy rate for truths in those studies is 73%, which is similar to the accuracy rates for lies, which is 72%. Accuracy rates for children do not seem to differ from accuracy rates for adults, further supporting that CBCA assessments are not restricted to children's statements.

To my knowledge, Ruby and Brigham (1998) are the only researchers to have

Table 4
Accuracy Rates

Authors	Age (years)	Event	Status	Assessment	Truth (%)	Lie (%)	Total (%)
Field studies							
Esplin et al. (1988)	3–15	Field	Victim	CBCA experts	100	100	100
Parker & Brown (2000)	Adult	Field	Victim	Decision rules	88	92	90
Laboratory studies							
Akehurst et al. (2001)	7–11/adult	Active	Na	Discriminant	73	67	70
Akehurst et al. (2001)	7–11	Active	Na	Discriminant			71
Akehurst et al. (2001)	Adult	Active	Na	Discriminant			90
Höfer et al. (1996)	Adult	Active	Na	Discriminant	70	73	71
Joffe & Yuille (1992)[a]	6–9	Active	Na	CBCA experts			71
Köhnken et al. (1995)	Adult	Video	Witness	Discriminant	89	81	85
Landry & Brigham (1992)	Adult	Memory	Victim	CBCA experts	75	35	55
Ruby & Brigham (1998)	Adult White	Memory	Victim	Discriminant	72	65	69
Ruby & Brigham (1998)	Adult Black	Memory	Victim	Discriminant	67	66	67
Santtila et al. (2000)	7–14 (total)	Memory	Victim	Regression	69	64	66
Sporer (1997)	Adult	Memory	Victim	Discriminant	70	60	65
Steller et al. (1988)	6–11	Memory	Victim	CBCA experts	78	62	72
Tye et al. (1999)	6–10	Active	Witness	Discriminant	75	100	89
Vrij, Edward, et al. (2000)	Adult	Video	Witness	Discriminant	65	80	73
Vrij, Kneller & Mann (2000)[b]	Adult	Video	Witness	Discriminant	53	80	67
Vrij, Kneller & Mann (2000)[b]	Adult	Video	Witness	CBCA experts	80	60	70
Vrij et al. (in press)	5–6	Active	Witness/suspect	Discriminant	71	64	69
Vrij et al. (in press)[b]	Adult	Active	Witness/suspect	Discriminant	67	75	71
Yuille (1988a)	6–9	Memory	Victim	CBCA experts	91	74	83
Zaparniuk et al. (1995)[c]	Adult	Video	Witness	Decision rule	80	77	78

Note. CBCA = Criteria-Based Content Analysis; Na = participants participated in an activity but were neither victims nor suspects.
[a]Lightly coached condition only. [b]Uninformed liars only. [c]Accuracy rates apply for the decision rule "Presence of Criteria 1–5, and any two of the remaining criteria."

examined the impact of ethnicity on the quality of statements (see also Vrij & Winkel, 1991, 1994, for ethnic differences in speech style). This issue merits attention in future studies given potential differences in narrative techniques between different cultures (Davies, 1994b; Phillips, 1993).

Validity Checklist

To date, Validity Checklist research has concentrated on the impact of three external factors included in the Validity Checklist (age of the interviewee, interviewer's style, and coaching of the interviewee) on CBCA scores.

Age of the Interviewee

Research has convincingly demonstrated that, as predicted, CBCA scores are positively correlated with age (Anson et al., 1993; Boychuk, 1991; Buck et al., 2002; Craig et al., 1999; Davies et al., 2000; Hershkowitz et al., 1997; Horowitz et al., 1997; Lamers-Winkelman & Buffing, 1996; Santtila et al., 2000; Vrij et al., 2002).[4]

Using statements from children in sexual abuse cases (with age varying between 4 and 12 years old), Anson et al. (1993) found that age was significantly correlated with logical structure, contextual embedding, description of interactions, reproduction of conversation, pardoning the perpetrator, and details characteristic of the offense. Interviews of allegedly sexually abused victims were also analyzed by Boychuk (1991), who compared CBCA scores of statements from children of different age groups (age varying from 4 to 16 years old) and found that descriptions of interactions, accounts of perpetrator's mental state, admitting lack of memory, and self-deprecation were more often present in the statements of older children (between 8 and 16 years old) than in the statements of younger children (between 4 and 7 years old; see Table 5). Statements of alleged sexual abuse victims (aged 2 to 11 years old) were analyzed by Lamers-Winkelman and Buffing (1996), and six criteria were found to be positively correlated with age: contextual embeddings, descriptions of interactions, reproduction of conversation, superfluous details, admitting lack of memory, and details characteristic of the offense.

Child sexual abuse interviews of children aged 2 to 14 years old were examined by Buck et al. (2002), who found that the total CBCA score and 13 of the 19 criteria were correlated with age. All 6 criteria that were not correlated (unusual details, accurately reported details misunderstood, attribution of perpetrator's mental state, raising doubts about one's own memory, self-deprecation, and pardoning the perpetrator) were present in less than 10% of the interviews.

In the only laboratory study to date examining age differences on individual CBCA criteria, Santtila et al. (2000) found that the youngest age group (7- to 8-year-olds) scored significantly lower on logical structure, quantity of details,

[4]Some studies did not obtain significant age effects (Akehurst et al., 2001; Tye et al., 1999). However, in Tye et al.'s (1999) study, children's ages were not balanced for true and false statements. The correlation between age and total CBCA score in Hershkowitz et al.'s (1997) study was only marginally significant ($p < .10$).

Table 5
Frequency of Occurrence of the CBCA Criteria (in Percentages) as a Function of Age

Authors	Age (years)	Event	Status	CBCA criterion 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Field studies																						
Boychuk (1991)	4–5	Field	Victim	80	80	67	73	20	40	33	27	53	20	27	40	0	53	13	13	0	20	67
Boychuk (1991)	6–7	Field	Victim	87	80	73	80	33	53	27	40	33	7	20	47	7	73	47	7	7	27	73
Boychuk (1991)	8–9	Field	Victim	93	87	87	73	60	60	60	53	40	7	20	47	13	67	60	13	13	27	80
Boychuk (1991)	10–12	Field	Victim	87	73	93	80	60	53	53	40	40	0	53	47	0	73	67	20	13	27	73
Boychuk (1991)	13–16	Field	Victim	100	80	93	87	67	73	73	27	40	7	20	73	13	80	80	7	27	40	87
Buck et al. (2002)	2–3	Field	Victim	65	0	55	65	15	30	5	0	5	10	0	10	0	30	5	0	5	0	25
Buck et al. (2002)	4	Field	Victim	61	0	72	83	22	39	6	6	17	17	17	11	6	32	11	0	0	0	33
Buck et al. (2002)	5–6	Field	Victim	74	9	78	91	30	35	22	9	26	0	30	22	0	48	44	0	9	0	35
Buck et al. (2002)	7–8	Field	Victim	82	6	88	88	24	53	12	24	35	12	41	41	12	65	30	0	6	0	41
Buck et al. (2002)	9–11	Field	Victim	93	36	93	100	50	64	29	7	64	0	36	71	14	43	43	0	7	14	71
Buck et al. (2002)	12–14	Field	Victim	100	42	100	100	50	75	17	8	33	8	33	42	0	67	67	8	0	0	75
Lamers-Winkelman & Buffing (1996)	2–3	Field	Victim	71	29	100	18	6	6	6	18	0	6	12	65	0	6	24	0	0	53	35
Lamers-Winkelman & Buffing (1996)	4–5	Field	Victim	72	41	100	21	13	21	5	18	08	0	26	69	13	15	39	0	0	64	62
Lamers-Winkelman & Buffing (1996)	6–8	Field	Victim	92	46	100	54	46	46	23	23	35	0	46	85	15	23	62	4	12	81	73
Lamers-Winkelman & Buffing (1996)	9–11	Field	Victim	95	62	100	71	67	62	33	38	52	10	91	100	29	43	76	14	0	67	100

Note. CBCA = Criteria-Based Content Analysis.

perpetrator's mental state, and spontaneous corrections compared with the oldest age group (aged 13–14 years old).[5]

Interview Style of the Interviewer

CBCA scores are also related to the interview style of the interviewer (Craig et al., 1999; Davies et al., 2000; Hershkowitz et al., 1997; Köhnken et al., 1995; Santtila et al., 2000; Steller & Wellershaus, 1996). For example, open-ended questions (Craig et al., 1999; Hershkowitz et al., 1997) and facilitators (nonsuggestive words of encouragement; Hershkowitz et al., 1997) yielded more CBCA criteria than other, more direct forms of questioning. Positive correlations between CBCA scores and verbal affirmations ("Yes, I see," etc.) and confirming comments (i.e., interviewer summarizing what the child has said) were found by Davies et al. (2000). Higher CBCA scores were received by statements obtained from interviewees who were interviewed with the cognitive interview technique, which facilitates the retrieval of information from memory, than from statements obtained using a standard interview technique in studies conducted by Köhnken et al. (1995) and Steller and Wellershaus (1996).

Coaching of the Interviewee

Finally, research has demonstrated that CBCA scores are related to coaching (Joffe & Yuille, 1992; Vrij et al., 2002; Vrij, Kneller, & Mann, 2000). For example, Vrij et al. (2002) gave their participants (10- to 11-year-olds, 14- to 15-year-olds, and undergraduates) some guidelines on how to tell a convincing story. In fact, they taught their participants several CBCA criteria. In a subsequent interview, these trained participants obtained higher CBCA scores than untrained participants.

Given that some external factors influence CBCA scores, do SVA experts take these factors into account when making their final judgments? Research about how the Validity Checklist is used in daily life is rare (but see Gumpert & Lindblad, 1999; Lamers-Winkelman, 1999; and Parker & Brown, 2000, for exceptions). However, several issues can be raised on the basis of the available psychological principles and research.

Some External Factors Might Be Hard to Detect

Some factors on the Validity Checklist might be difficult to identify. SVA experts look for evidence that an adult might have coached the child to enhance the perceived credibility of statements. For example, in a bitter divorce settlement, one parent might use the dubious tactic of falsely exposing his or her ex-spouse as a child abuser to enhance the chances of winning custody of the children. In their experiments, Joffe and Yuille (1992) and Vrij, Kneller, and Mann (2000) coached some participants and asked a trained CBCA expert to examine the statements of all participants (truth tellers, coached liars, and uncoached liars).

[5]The remaining researchers who examined age differences (Anson et al., 1993; Craig et al., 1999; Davies et al., 2000; Hershkowitz et al., 1997; Horowitz et al., 1997; Vrij et al., 2002) did not report age differences for individual CBCA criteria.

The CBCA experts did not notice that some participants had been coached and did not discriminate successfully between truth tellers and coached liars. Perhaps causing even more concern, in Vrij, Kneller, and Mann's study, the CBCA experts still could not indicate which statements belonged to the coached liars even after they had been informed that some of the participants had been coached.

Difficulty in Measuring Some External Factors

At least one external factor, susceptibility to suggestion (Criterion 3; Steller, 1989), is difficult to measure. Some witnesses are more prone to suggestions made by interviewers than others, and a suggestible child might be more inclined to provide information that confirms the interviewer's expectations but is, in fact, inaccurate. Yuille (1988b) therefore recommended asking the witness at the end of the interview a few leading questions to assess susceptibility to suggestion. He recommended asking some questions about peripheral rather than central information as asking leading questions might distort the interviewee's memory and therefore harm the case (Loftus & Palmer, 1974). The fact that questions can be asked only about peripheral information is problematic as it may say little about the witness's suggestibility regarding core issues of his or her statement. Children show more resistance to suggestibility for central parts than peripheral parts of an event (Goodman, Rudy, Bottoms, & Aman, 1990). They are also more resistant to suggestibility for stressful events, most likely the central events, than for events that are less stressful, most likely to be peripheral events (Davies, 1991). Thus, if an interviewee yields to a leading question about a peripheral part of the event, this does not imply that he or she is not resistant to suggestion when more important incidents are discussed. This criterion also seems to assume that suggestion is more the result of individual differences than of circumstances. This may not be a valid assumption (Milne & Bull, 1999).

Some Relevant External Factors Are Not in the Validity Checklist

Some factors that influence CBCA scores are not present in the Validity Checklist. Research has shown that CBCA scores are related to verbal and social skills (Santtila et al., 2000; Vrij, Edward, & Bull, 2001b; Vrij et al., 2002). For example, Santtila et al. (2000) found a positive correlation between CBCA scores and verbal ability, and Vrij et al. (2002) found that CBCA scores were, in some age groups, positively correlated with social adroitness and self-monitoring and negatively correlated with social anxiety. However, this is not taken into account by SVA experts when they rely on the Validity Checklist. Although the Validity Checklist might look like a complete list of external factors, in fact, it is not. Obviously, other external factors, presently unknown, might have an impact on CBCA scores as well. For example, children required to give statements might well have psychological disorders, such as depression or attentional problems, and it is unclear what impact this might have on their CBCA scores.

Obviously, the issue of difficulty in measuring external factors discussed above applies to these concepts as well. For example, how should issues such as social adroitness and self-monitoring be assessed in the individual case?

Some False Allegations Might Be Virtually Impossible to Detect

There are at least three types of false allegations that are very hard to detect. First, the situation in which someone has been sexually abused but misidentifies the perpetrator and instead accuses an innocent suspect of being the culprit. Such a statement might be rich in detail and might obtain a high CBCA score, as most of the account is true. The accusation is nevertheless false. In general, a CBCA assumption is that a statement is either totally truthful or totally fabricated, and there is no procedure to distinguish between experienced and nonexperienced portions within the same account. Second, sometimes people, both adults and children, are confused about what they have actually experienced and what they have only imagined. Research has demonstrated that imagined narratives can be internally coherent and detailed (Ceci, Huffman, Smith, & Loftus, 1994; Ceci, Loftus, Leichtman, & Bruck, 1994; Porter, Yuille, & Lehman, 1999). They are therefore likely to obtain high CBCA scores and to be judged as truthful. Third, the coaching studies described above (Joffe & Yuille, 1992; Vrij et al., in press; Vrij, Kneller, & Mann, 2000) have revealed that well-prepared lies that include many CBCA criteria are difficult to detect.

Justification of Some External Factors on the Validity Checklist

It is possible to question the justification of some of the external factors listed on the Validity Checklist, such as Criterion 2, inappropriateness of affect (Steller, 1989); Criterion 10, inconsistency with other statements (Steller, 1989); Criterion 9, consistency with the law of nature (Steller, 1989); and Criterion 11, consistency with other evidence (Steller, 1989).

Criterion 2 refers to whether the child displays an absence of affect or inappropriate affect during the interview (Raskin & Esplin, 1991b). It suggests that if a child reports details of abuse without showing any signs of emotion or showing inappropriate signs of emotion, the story might be less trustworthy. This view on emotional displays is too rigid as the notion of an appropriate affect does not exist. Research with rape victims has distinguished two basic styles of self-presentation: an expressed style in which the victim displays distress that is clearly visible to outsiders and a more controlled, numbed style whereby cues of distress are not clearly visible (Burgess, 1985; Burgess & Holmstrom, 1974). Although the styles represent a personality factor and are not related to deceit (Littmann & Szewczyk, 1983), they have a differential impact on the perceived credibility of victims, and emotional victims are more readily believed than victims who report their experience in a more controlled manner (Baldry, Winkel, & Enthoven, 1997; Kaufmann, Drevland, Wessel, Overskeid, & Magnussen, 2003; Vrij & Fisher, 1997; Winkel & Koppelaar, 1991). Given that inappropriate affect does not exist and that people tend to draw conclusions on the basis of the displayed affect that are not always correct, it is unfortunate to encourage the evaluator to pay attention to such affect.

Criterion 10 deals with inconsistencies between different statements from the same witness. It suggests that one statement may in fact be fabricated when interviewees contradict themselves in two different statements. This belief might be incorrect. In their research with adult participants, Granhag and Strömwall (1999, 2002) have demonstrated that inconsistency between different statements

is not a valid indicator of deception. In their review of child research, Fivush, Peterson, and Schwarzmueller (2002) also concluded that inconsistency, in itself, is not an indication of inaccuracy. Neither is it the case, as these authors pointed out, that consistency necessarily means accuracy. Moreover, Fivush's own research, reported in Fivush et al., has demonstrated that children's narratives naturally change across recall occasions, which is largely due to differences in interviewers across interviews and type of questions asked. Finally, judging whether a statement is consistent or not might be more difficult than it initially appears. Judges often do not agree among themselves whether a statement is consistent with a previous statement or not, according to research by Granhag and Strömwall (2001a, 2001b).

Criteria 9 and 11 deal with the realism of the statement. Dalenberg, Hyland, and Cuevas (2002) reported that for a small group of children who made initial allegations of abuse and for whom there was a gold standard of proof that abuse had occurred (the allegations were supported by confessions, and the injuries were judged medically consistent with the allegations), bizarre and improbable material was included in their statements (reference to fantasy figures, impossible or extremely implausible features of the story, and descriptions of extreme abusive acts that should have been [but were not] supported by external evidence if they had genuinely occurred). How would SVA experts handle such allegations? Such statements clearly contain unrealistic elements, and there is therefore a risk that such allegations would be considered as untrue on the basis of the Validity Checklist assessment.

Difficulty in Determining the Exact Impact of an External Factor

Even when an SVA expert knows that an external factor that appears on the Validity Checklist is present, it is still difficult to determine the exact impact of that factor on CBCA scores. In a field study, raters were instructed to take the age of the child into account (Lamers-Winkelman & Buffing, 1996). Nevertheless, six criteria positively correlated with age. Alternatively, CBCA raters may rate all statements in the same way, regardless of the age of the interviewee, but may apply different decision rules afterward for different age groups: For example, in younger children, a statement is likely to be truthful if five criteria are present; in older children, at least eight criteria should be present; and so on. However, applying decision rules is impossible because it is unknown what age-related cutoff marks should be used. Given these difficulties in identifying the relevant external factors and in examining the exact impact of these factors on CBCA scores, it is clear that the Validity Checklist procedure is more subjective and less formalized than the CBCA procedure (Steller, 1989; Steller & Köhnken, 1989). It is therefore not surprising that if two experts disagree about the truthfulness of a statement in German criminal cases, they often disagree about the likely impact of some external factors on that statement (G. Köhnken, personal communication, 1997).

In their field study concerning the use of the Validity Checklist in Sweden, Gumpert and Lindblad (1999) showed that different experts sometimes drew different conclusions about the impact of external factors on children's statements. It is therefore advisable that, in applied settings, not one but at least two

evaluators assess a case independent of each other. At present, this is not common practice.[6]

No Guidelines to Determine the Weighting of CBCA Criteria

Steller and Köhnken (1989) noted that some criteria might be of more value in assessing truthfulness than others. For example, the presence of accurately reported but misunderstood details in a statement (Criterion 10), such as a child who describes the adult's sexual behavior but attributes it to a sneeze or to pain, is apparently more significant than the fact that the child describes where the alleged sexual encounter took place (Criterion 4).[7] However, no guidance is given in the SVA procedure regarding how a different weighting system should be applied, leaving this to the interpretation of the individual expert.

The Validity Checklist Might Be Improperly Used

Gumpert and Lindblad's (1999) field study regarding the use of the Validity Checklist by SVA experts in Sweden revealed that these experts might have used this list incorrectly. First, although SVA experts sometimes highlighted the influence of external factors on children's statements in general, they did not always discuss how this factor might have influenced the statement of the particular child they were asked to assess. Second, although experts sometimes indicated possible external influence on statements, they tended to rely on the CBCA outcome and tended to judge high-quality statements as truthful and low-quality statements as fabricated. Although Gumpert and Lindblad examined only a limited number of cases and, so, to draw convincing conclusions would perhaps be premature, their findings cause concern. They implied that SVA decisions are not likely to be more accurate than CBCA assessments as the final decision based on CBCA outcomes, together with the Validity Checklist procedure, often is the same as the decision based on CBCA outcomes alone. They also implied that interviewees who naturally produce low-quality statements and

---

[6]The problem CBCA/SVA evaluators have to deal with—that a witness's response is influenced not just by the veracity of a statement but also by external factors—is not unique to SVA assessments but happens in physiological and nonverbal lie detection as well. Those latter lie-detection techniques attempt to resolve the issue by introducing a baseline response that is a typical, natural response of the interviewee that the lie detector knows to be a truthful response and that is provided in circumstances similar to the response under investigation. They then compare the baseline response with the response under investigation, and because, in that situation, the impact of external factors on both responses is assumed to be the same, differences between the two responses may indicate deception. However, the method is complex as creating a good baseline is often problematic (Vrij, 2002b).

[7]Horowitz (1991) pointed out that it is dangerous to form an impression about the veracity of a statement on the basis of a child's knowledge about sexual matters as there are no age norms for such knowledge (Jones & McQuiston, 1989). Moreover, Gordon, Schroeder, and Abrams (1990), who compared abused children with a matched sample of nonabused children on sexual knowledge, found no differences between these two groups. Despite this, many professionals consider so-called age-inappropriate sexual knowledge an important indicator of sexual abuse (Conte, Sorenson, Fogarty, & Rosa, 1991). As mentioned before, there is not much empirical evidence to support the idea that Criterion 10 occurs more frequently in truthful responses, perhaps because this criterion is seldom present in statements at all.

therefore are likely to obtain low CBCA scores (i.e., young children, interviewees with poor verbal skills, etc.) might well be in a disadvantageous position.

Legal Implications

What are the implications of these findings for the use of CBCA/SVA assessments as scientific evidence in legal systems? A possible way to answer this question is by examining to what extent CBCA/SVA assessments meet the criteria that are required for admitting expert scientific evidence in criminal courts. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), the United States Supreme Court promulgated a set of guidelines for admitting expert scientific evidence in the (American) federal courts. The following guidelines were provided by the Supreme Court and reported and discussed by Honts (1994): (a) Is the scientific hypothesis testable, (b) has the proposition been tested, (c) is there a known error rate, (d) has the hypothesis and/or technique been subjected to peer review and publication, and (e) is the theory on which the hypothesis and/or technique is based generally accepted in the appropriate scientific community?

The answer to the first question—Is the scientific hypothesis testable?—is yes. The Undeutsch hypothesis can be tested in scientific research, although, as this review has revealed, this is not an easy task. The Undeutsch hypothesis can easily be tested in experimental laboratory-based research, but the findings might not be ecologically valid given the artificial nature of such studies. Testing the Undeutsch hypothesis in field studies is possible in principle; however, in practice, it is difficult to establish the truth or falsity of statements beyond doubt.

The answer to the second question—Has the proposition been tested?—is also suggested to be affirmative; however, most of the studies indicating this have been experimental laboratory studies, and in most studies, adults rather than children participated. There are very few properly conducted field studies testing the Undeutsch hypothesis. In general, the available studies provide empirical support for the Undeutsch hypothesis. In 11 out of 12 studies in which a total CBCA score was calculated, the CBCA score was significantly higher for truth tellers than for liars, which supports the Undeutsch hypothesis. When the individual criteria are taken into account, the criteria with the strongest support (Criteria 2, 3, 4, and 6) are all part of the cognitive component of the Undeutsch hypothesis (Criteria 1–13). Support for the motivational component of the hypothesis (Criteria 14–18) is generally weak.

The answer to the third question—Is there a known error rate?—is no. Clearly, there is a known error rate of CBCA judgments made in experimental laboratory research, which is approximately 30% for both detecting truths and detecting lies. However, of particular interest here is the error rate of SVA judgments in field studies. A properly conducted study examining this issue has not been published to date. As long as the error rate in field studies is unknown, there is no better alternative than to use the known error rate in CBCA laboratory studies. This error rate, around 30%, is probably not an unreasonable estimate for the accuracy of SVA judgments. There are reasons to believe that truth–lie assessments in real-life situations are as difficult as or even more difficult than truth–lie assessments in experimental laboratory studies. Research, reviewed in the Validity Checklist section of this review, has demonstrated that CBCA scores

are affected not only by the veracity of the statement but also by other factors, such as age, verbal ability, and social skills of the interviewee and the interview style of the interviewer. In the Validity Checklist section, it has also been argued that it is difficult in real-life situations to indicate which external factors might have influenced the quality of the statement. Some external factors (such as coaching of the interviewee) are difficult to detect, and interviewees who come to know about the method might therefore dupe evaluators. Other factors (such as social skills of the interviewee) are not included in the Validity Checklist and are therefore likely to be ignored by evaluators. Further factors (e.g., whether the interviewee was suggestible during the interview) are difficult to measure. Finally, I have raised some concerns about the appropriateness of some factors (such as looking for consistency between statements).

Moreover, it is difficult to determine the exact impact of these external factors on a particular statement. For example, even in studies in which raters were instructed to take the child's age into account (Lamers-Winkelman & Buffing, 1996), CBCA scores still correlated with age. In one of the very few studies regarding the Validity Checklist, Gumpert and Lindblad (1999) found that SVA experts had the tendency to rely heavily on the CBCA outcomes and that a high-quality statement was often considered to be true and a low-quality statement was often considered to be false. The combined findings of Lamers-Winkelman and Buffing (1996) and Gumpert and Lindblad (1999) suggest that young interviewees, who naturally produce low CBCA scores, are in a disadvantageous position. Other interviewees who naturally produce low-quality statements (such as interviewees with poor verbal skills, socially inept interviewees, etc.) might be in a similarly disadvantageous position. Finally, a further complication in making SVA assessments is that some false allegations (i.e., false narratives that contain many true elements, false memories, well-prepared lies) are difficult to detect.

In summary, although the error rates for SVA assessments in real-life cases are unknown, incorrect decisions are likely to occur given the numerous difficulties associated with making SVA assessments. If one takes the known error rate of 30% as a guideline, than it is clear that SVA evaluators are not able to present the accuracy of their SVA assessments as being beyond reasonable doubt, which is the standard of proof often set in criminal courts. In other words, SVA assessments are not accurate enough to be presented as scientific evidence in criminal courts.

The answer to the fourth question—Has the hypothesis and/or technique been subjected to peer review and publication?—is again yes. A growing number of CBCA studies have now been published in peer reviewed journals, although, again, most studies were laboratory-based studies in which the participants were often adults rather than children.

The answer to the fifth and final question—Is the theory on which the hypothesis and/or technique is based generally accepted in the appropriate scientific community?—is probably no. As already mentioned in the introductory section, several authors have expressed serious doubts about the method (Brigham, 1999; Davies, 2001; Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997; Rassin, 1999; Ruby & Brigham, 1997; Wells & Loftus, 1991). However, a proper survey, similar to the one in which scientific opinion concern-

ing the polygraph was examined (Iacono & Lykken, 1997), has not been published to date.

Conclusions

SVA evaluations do not meet the *Daubert* (1993) guidelines for admitting expert scientific evidence in criminal courts. The two main reasons are that the error rate is too high and that the method is not undisputed in the relevant scientific community. Regarding the high error rate, SVA evaluators might challenge the claim that the error rate is around 30% as this is the known error rate for CBCA assessments made in laboratory studies rather than the error rate for SVA evaluations made in real-life situations. However, those SVA evaluators should realize that in case CBCA error rates should be negated, all that could then be concluded is that the error rate is unknown, an outcome that does not meet the *Daubert* guideline either.

At present, SVA evaluations are accepted as evidence in criminal courts in several countries. In those countries, at the very least, SVA experts should present the problems and limitations of SVA assessments in court so that judges, jurors, prosecutors, and solicitors can make an informed decision about the validity of SVA decisions. In addition, although the interrater agreement rates between CBCA judges are generally adequate, they are not perfect and are likely to be higher than the interrater agreement rates regarding the Validity Checklist. This all clearly makes conducting SVA judgments a subjective exercise, and therefore, more than one expert should judge each statement to establish interrater reliability between evaluators.

However, true and fabricated stories can be detected above the level of chance with CBCA/SVA assessments in both children and adults and in contexts other than sexual abuse incidents, which makes such assessments a valuable tool for police investigations. They might be useful, for example, in the initial stage of investigation for forming rough indications of the veracity of various statements in cases in which police detectives have different opinions about the veracity of a statement. Thorough training in how to conduct CBCA/SVA assessments is probably desirable given the erratic effects obtained in previous studies in which trainees were exposed to less comprehensive training programs.

References

References marked with an asterisk indicate studies included in the literature review.

*Akehurst, L., Bull, R., & Vrij, A. (1998, September). *Training British police officers, social workers and students to detect deception in children using Criteria-Based Content Analysis*. Paper presented at the 8th European Conference of Psychology and Law, Krakow, Poland.

*Akehurst, L., Köhnken, G., & Höfer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6,* 65–83.

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

Arntzen, F. (1982). Die Situation der Forensischen Aussagenpsychologie in der Bundesrepubliek Deutschland [The state of forensic psychology in the Federal Republic of Germany]. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 107–120). Deventer, the Netherlands: Kluwer.

Baldry, A. C., Winkel, F. W., & Enthoven, D. S. (1997). Paralinguistic and nonverbal triggers of biased credibility assessments of rape victims in Dutch police officers: An experimental study of "nonevidentiary" bias. In S. Redondo, V. Garrido, J. Perze, & R. Barbaret (Eds.), *Advances in psychology and law* (pp. 163–174). Berlin, Germany: Walter de Gruyter.

Berliner, L., & Conte, J. R. (1993). Sexual abuse evaluations: Conceptual and empirical obstacles. *Child Abuse and Neglect, 17,* 111–125.

*Boychuk, T. (1991). *Criteria-Based Content Analysis of children's statements about sexual abuse: A field-based validation study*. Unpublished doctoral dissertation, Arizona State University.

Bradford, R. (1994). Developing an objective approach to assessing allegations of sexual abuse. *Child Abuse Review, 3,* 93–101.

Brigham, J. C. (1999). What is forensic psychology, anyway? *Law and Human Behavior, 23,* 273–298.

*Buck, J. A., Warren, A. R., Betman, S., & Brigham, J. C. (2002). Age differences in Criteria-Based Content Analysis scores in typical child sexual abuse interviews. *Applied Developmental Psychology, 23,* 267–283.

Bull, R. (1992). Obtaining evidence expertly: The reliability of interviews with child witnesses. *Expert Evidence: The International Digest of Human Behaviour Science and Law, 1,* 3–36.

Bull, R. (1995). Innovative techniques for the questioning of child witnesses, especially those who are young and those with learning disability. In M. Zaragoza (Ed.), *Memory and testimony in the child witness* (pp. 179–195). Thousand Oaks, CA: Sage.

Bull, R. (1998). Obtaining information from child witnesses. In A. Memon, A. Vrij, & R. Bull, *Psychology and law: Truthfulness, accuracy and credibility* (pp. 188–210). Maidenhead, England: McGraw-Hill.

Burgess, A. W. (1985). *Rape and sexual assault: A research book*. London: Garland.

Burgess, A. W., & Holmstrom, L. L. (1974). *Rape: Victims of crisis*. Bowie, MD: Brady.

Bybee, D., & Mowbray, C. T. (1993). An analysis of allegations of sexual abuse in a multi-victim day-care center case. *Child Abuse and Neglect, 17,* 767–783.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom*. Washington, DC: American Psychological Association.

Ceci, S. J., Huffman, M. L., Smith, E., & Loftus, E. F. (1994). Repeatedly thinking about a non-event. *Consciousness and Cognition, 3,* 388–407.

Ceci, S. J., Loftus, E. F., Leichtman, M. D., & Bruck, M. (1994). The possible role of source misattributions in the creation of false beliefs among preschoolers. *International Journal of Clinical and Experimental Hypnosis, 17,* 304–320.

*Colwell, K., Hiscock, C. K., & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology, 16,* 287–300.

Conte, J. R., Sorenson, E., Fogarty, L., & Rosa, J. D. (1991). Evaluating children's reports of sexual abuse: Results from a survey of professionals. *Journal of Orthopsychiatry, 61,* 428–437.

*Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. H. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. *Applied Developmental Science, 3,* 77–85.

Dalenberg, C. J., Hyland, K. Z., & Cuevas, C. A. (2002). Sources of fantastic elements in allegations of abuse by adults and children. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 185–204). Mahwah, NJ: Erlbaum.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993).

Davies, G. M. (1991). Research on children's testimony: Implications for interviewing

practice. In C. R. Hollin & K. Howells (Eds.), *Clinical approaches to sex offenders and their victims* (pp. 177–191). New York: Wiley.

Davies, G. M. (1994a). Children's testimony: Research findings and police implications. *Psychology, Crime, and Law, 1,* 175–180.

Davies, G. M. (1994b). Statement validity analysis: An art or a science? Commentary on Bradford. *Child Abuse Review, 3,* 104–106.

Davies, G. M. (2001). Is it possible to discriminate true from false memories? In G. M. Davies & T. Dalgleish (Eds.), *Recovered memories: Seeking the middle ground* (pp. 153–176). Chichester, England: Wiley.

*Davies, G. M., Westcott, H. L., & Horan, N. (2000). The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. *Psychology, Crime, and Law, 6,* 81–97.

DePaulo, B. M., Stone, J. L., & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. R. Schenkler (Ed.), *The self and social life* (pp. 323–370). New York: McGraw-Hill.

Doris, J. (1994). Commentary on Criteria-Based Content Analysis. *Journal of Applied Developmental Psychology, 15,* 281–285.

*Esplin, P. W., Boychuk, T., & Raskin, D. C. (1988, June). *A field validity study of Criteria-Based Content Analysis of children's statements in sexual abuse cases*. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Fivush, R., Haden, C., & Adam, S. (1995). Structure and coherence of preschoolers' personal narratives over time: Implications for childhood amnesia. *Journal of Experimental Child Psychology, 60,* 32–56.

Fivush, R., Peterson, C., & Schwarzmueller, A. (2002). Questions and answers: The credibility of child witnesses in the context of specific questioning techniques. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 331–354). Mahwah, NJ: Erlbaum.

Flavell, J. H., Botkin, P. T., Fry, C. K., Wright, J. C., & Jarvis, P. T. (1968). *The development of role-taking and communication skills in children*. New York: Wiley.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.

Goodman, G. S., Rudy, L., Bottoms, B., & Aman, C. (1990). Children's concerns and memory: Issues of ecological validity in the study of children's eyewitness testimony. In R. Fivush & J. Hudson (Eds.), *Knowing and remembering in young children* (pp. 249–284). New York: Cambridge University Press.

Gordon, B. N., Schroeder, C. S., & Abrams, J. M. (1990). Children's knowledge of sexuality: A comparison of sexually abused and nonabused children. *American Journal of Orthopsychiatry, 60,* 250–257.

Granhag, P. A., & Strömwall, L. A. (1999). Repeated interrogations: Stretching the deception detection paradigm. *Expert Evidence: The International Journal of Behavioural Sciences in Legal Contexts, 7,* 163–174.

Granhag, P. A., & Strömwall, L. A. (2001a). Deception detection: Examining the consistency heuristic. In C. M. Breur, M. M. Kommer, J. F. Nijboer, & J. M. Reijntjes (Eds.), *New trends in criminal investigation and evidence* (Vol. 2, pp. 309–321). Antwerpen, Belgium: Intresentia.

Granhag, P. A., & Strömwall, L. A. (2001b). Deception detection: Interrogators' and observers' decoding of consecutive statements. *Journal of Psychology, 135,* 603–620.

Granhag, P. A., & Strömwall, L. A. (2002). Repeated interrogations: Verbal and nonverbal cues to deception. *Applied Cognitive Psychology, 16,* 243–257.

Gumpert, C. H., & Lindblad, F. (1999). Expert testimony on child sexual abuse: A qualitative study of the Swedish approach to statement analysis. *Expert Evidence, 7,* 279–314.

Gumpert, C. H., Lindblad, F., & Grann, M. (2002a). The quality of written expert testimony in alleged child sexual abuse: An empirical study. *Psychology, Crime, and Law, 8,* 77–92.

Gumpert, C. H., Lindblad, F., & Grann, M. (2002b). A systematic approach to quality assessment of expert testimony in cases of alleged child sexual abuse. *Psychology, Crime, and Law, 8,* 59–75.

Hershkowitz, I. (1999). The dynamics of interviews yielding plausible and implausible allegations of child sexual abuse. *Applied Developmental Science, 3,* 28–33.

Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology, 6,* 49–63.

*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.

*Höfer, E., Akehurst, L., & Metzger, G. (1996, August). *Reality monitoring: A chance for further development of CBCA?* Paper presented at the annual meeting of the European Association on Psychology and Law, Siena, Italy.

Honts, C. R. (1994). Assessing children's credibility: Scientific and legal issues in 1994. *North Dakota Law Review, 70,* 879–903.

Horowitz, S. W. (1991). Empirical support for Statement Validity Assessment. *Behavioral Assessment, 13,* 293–313.

*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.

Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuk, T. D., Reiter-Lavery, L., & Krispin, O. (1996). Establishing ground truth in studies of child sexual abuse. *Expert Evidence: The International Digest of Human Behaviour Science and Law, 4,* 42–52.

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology, 82,* 426–433.

*Joffe, R., & Yuille, J. C. (1992, May). *Criteria-Based Content Analysis: An experimental investigation*. Paper presented at the NATO Advanced Study Institute on the Child Witness in Context: Cognitive, Social and Legal Perspectives, Lucca, Italy.

Jones, D. P. H., & McQuiston, M. (1989). *Interviewing the sexually abused child*. London: Gaskell.

Kaufmann, G., Drevland, G. C., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: Displayed emotions and witness credibility. *Applied Cognitive Psychology, 17,* 21–34.

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour, 2,* 1–17.

Köhnken, G. (1989). Behavioral correlates of statement credibility: Theories, paradigms and results. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 271–289). New York: Springer-Verlag.

Köhnken, G. (1996). Social psychology and the law. In G. R. Semin & K. Fiedler (Eds.), *Applied social psychology* (pp. 257–282). London: Sage.

Köhnken, G. (1999, July). *Statement Validity Assessment*. Paper presented at the preconference program of applied courses assessing credibility organized by the European Association of Psychology and Law, Dublin, Ireland.

Köhnken, G. (2002). A German perspective on children's testimony. In H. L. Westcott, G. M. Davies, & R. H. C. Bull (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 233–244). Chichester, England: Wiley.

*Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive

interview and the assessment of the credibility of adult's statements. *Journal of Applied Psychology, 80,* 671–684.

Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in the German procedural system. In G. Davies & J. Drinkwater (Eds.), *The child witness: Do the courts abuse children?* (pp. 37–45). Leicester, England: British Psychological Society.

Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1994). Factors influencing the reliability and validity of statements made by young victims of sexual maltreatment. *Journal of Applied Developmental Psychology, 15,* 255–280.

Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1998). Conducting investigative interviews of alleged sexual abuse victims. *Child Abuse and Neglect, 22,* 813–823.

Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., & Orbach, Y. (1997). Assessing the credibility of children's allegations of sexual abuse: A survey of recent research. *Learning and Individual Differences, 9,* 175–194.

*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.

Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., & Esplin, P. W. (1999). Forensic interviews of children. In R. Bull & A. Memon (Eds.), *The psychology of interviewing: A handbook* (pp. 253–277). Chichester, England: Wiley.

Lamers-Winkelman, F. (1995). *Seksueel misbruik van jonge kinderen: Een onderzoek naar signalen en signaleren, en naar ondervragen en vertellen inzake seksueel misbruik* [Sexual abuse of young children]. Amsterdam: VU Uitgeverij.

Lamers-Winkelman, F. (1999). Statement Validity Analysis: Its application to a sample of Dutch children who may have been sexually abused. *Journal of Aggression, Maltreatment and Trauma, 2,* 59–81.

*Lamers-Winkelman, F., & Buffing, F. (1996). Children's testimony in the Netherlands: A study of Statement Validity Analysis. In B. L. Bottoms & G. S. Goodman (1996), *International perspectives on child abuse and children's testimony* (pp. 45–62). Thousand Oaks, CA: Sage.

*Landry, K., & Brigham, J. C. (1992). The effect of training in Criteria-Based Content Analysis on the ability to detect deception in adults. *Law and Human Behavior, 16,* 663–675.

Littmann, E., & Szewczyk, H. (1983). Zu einigen Kriterien und Ergebnissen forensisch-psychologischer Glaubwürdigkeitsbegutachtung von sexuell misbrauchten Kindern und Jugendlichen [Criteria and results of forensic-psychological reports concerning the credibility of sexually abused children and adolescents]. *Forensia, 4,* 55–72.

Loftus, E. F., & Palmer, J. C. (1974). Reconstructions of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.

Maxwell, A. E. (1977). Coefficients of agreement between observers and their interpretation. *British Journal of Psychiatry, 130,* 79–83.

Memon, A., & Bull, R. (1999). *Handbook of the psychology of interviewing*. Chichester, England: Wiley.

Milne, R., & Bull, R. (1999). *Investigative interviewing: Psychology and practice*. Chichester, England: Wiley.

Moston, S. J., Stephenson, G. M., & Williamson, T. M. (1992). The effects of case characteristics on suspect behaviour during police questioning. *British Journal of Criminology, 32,* 23–39.

*Orbach, Y., & Lamb, M. E. (1999). Assessing the accuracy of a child's account of sexual abuse: A case study. *Child Abuse and Neglect, 23,* 91–98.

*Parker, A. D., & Brown, J. (2000). Detection of deception: Statement Validity Analysis

as a means of determining truthfulness or falsity of rape allegations. *Legal and Criminological Psychology, 5,* 237–259.

Pezdek, K., & Taylor, J. (2000). Discriminating between accounts of true and false events. In D. F. Bjorklund (Ed.), *Research and theory in false memory creation in children and adults* (pp. 69–91). Mahwah, NJ: Erlbaum.

Phillips, M. (1993). Investigative interviewing: Issues of race and culture. In W. Stainton-Rogers & M. Worrel (Eds.), *Investigative interviewing with children* (pp. 50–55). Milton Keynes, England: Open University Press.

*Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior, 20,* 443–459.

*Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior, 23,* 517–537.

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. 153–165). 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, 13,* 265–291.

Raskin, D. C., & Steller, M. (1989). Assessing the credibility of allegations of child sexual abuse: Polygraph examinations and statement analysis. In H. Wegener, F. Losel, & J. Haisch (Eds.), *Criminal behavior and the justice system* (pp. 290–302). New York: Springer.

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

Rassin, E. (1999). Criteria-Based Content Analysis: The less scientific road to truth. *Expert Evidence, 7,* 265–278.

Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the Criteria-Based Content Analysis technique in distinguishing between truthful and fabricated allegations. *Psychology, Public Policy, and Law, 3,* 705–737.

*Ruby, C. L., & Brigham, J. C. (1998). Can Criteria-Based Content Analysis distinguish between true and false statements of African-American speakers? *Law and Human Behavior, 22,* 369–388.

*Santtila, P., Roppola, H., Runtti, M., & Niemi, P. (2000). Assessment of child witness statements using Criteria-Based Content Analysis (CBCA): The effects of age, verbal ability, and interviewer's emotional style. *Psychology, Crime, and Law, 6,* 159–179.

Spitznagel, E. L., & Helzer, J. E. (1985). A proposed solution to the base rate problem in the kappa statistic. *Archives of General Psychiatry, 42,* 725–728.

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

Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135–154). Deventer, the Netherlands: Kluwer.

Steller, M., & Boychuk, T. (1992). Children as witnesses in sexual abuse cases: Investigative interview and assessment techniques. In H. Dent & R. Flin (Eds.), *Children as witnesses* (pp. 47–73). New York, NJ: John Wiley & Sons.

Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). New York: Springer-Verlag.

*Steller, M., & Wellershaus, P. (1996). Information enhancement and credibility assess-

ment of child statements: The impact of the cognitive interview on Criteria-Based Content Analysis. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), *Psychology, law, and criminal justice: International developments in research and practice* (pp. 118–127). Berlin, Germany: Walter de Gruyter.

*Steller, M., Wellershaus, P., & Wolf, T. (1988, June). *Empirical validation of Criteria-Based Content Analysis*. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Sternberg, K. J., Lamb, M. E., Esplin, P. W., Orbach, Y., & Hershkowitz, I. (2002). Using a structured interview protocol to improve the quality of investigative interviews. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 409–436). Mahwah, NJ: Erlbaum.

Strömwall, L. A., & Granhag, P. A. (2003). Hoe to detect deception? Arresting the beliefs of police officers, prosecutors and judges. *Psychology, Crime, and Law, 9,* 19–36.

Trankell, A. (1972). *Reliability of evidence.* Stockholm, Beckmans.

Tully, B. (1999). Statement validation. In D. Canter & L. Alison (Eds.), *Interviewing and deception* (pp. 83–104). Aldershot, England: Darmouth.

*Tye, M. C., Amato, S. L., Honts, C. R., Kevitt, 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. (1967). Beurteilung der Glaubhaftigkeit von Aussagen [Veracity assessment of statements]. In U. Undeutsch (Ed.), *Handbuch der Psychologie: Vol. 11. Forensische Psychologie* (pp. 26–181). Göttingen, Germany: Hogrefe.

Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 27–56). Deventer, the Netherlands: Kluwer.

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

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and its implications for professional practice*. Chichester, England: Wiley.

Vrij, A. (2002a). Deception in children: A literature review and implications for children's testimony. In H. L. Westcott, G. M. Davies, & R. H. C. Bull (Eds.), *Children's testimony* (pp. 175–194). Chichester, England: Wiley.

Vrij, A. (2002b). Telling and detecting lies. In N. Brace & H. Westcott (Eds.), *Applying psychology* (pp. 179–242). Milton Keynes, England: Open University.

*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). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research, 30,* 8–41.

*Vrij, A., Edward, K., & Bull, R. (2001a). People's insight into their own behaviour and speech content while lying. *British Journal of Psychology, 92,* 373–389.

*Vrij, A., Edward, K., & Bull, R. (2001b). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin, 27,* 899–909.

*Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24,* 239–263.

Vrij, A., & Fisher, A. (1997). The role of displays of emotions and ethnicity in judgements of rape victims. *International Review of Victimology, 4,* 255–265.

*Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime, and Law, 5,* 203–315.

*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.

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior, 15,* 169–184.

Vrij, A., & Winkel, F. W. (1994). Perceptual distortions in cross-cultural interrogations: The impact of skin color, accent, speech style and spoken fluency on impression formation. *Journal of Cross-Cultural Psychology, 25,* 284–295.

Walker, A. G., & Warren, A. R. (1995). The language of the child abuse interview: Asking the questions, understanding the answers. In T. Ney (Ed.), *True and false allegations in child sexual abuse: Assessment and case management* (pp. 153–162). New York: Brunner-Mazel.

Wegener, H. (1989). The present state of statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 121–134). Dordrecht, the Netherlands: Kluwer.

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

Winkel, F. W., & Koppelaar, L. (1991). Rape victims' style of self-presentation and secondary victimization by the environment. *Journal of Interpersonal Violence, 6,* 29–40.

*Winkel, F. W., & Vrij, A. (1995). Verklaringen van kinderen in interviews: Een experimenteel onderzoek naar de diagnostische waarde van Criteria-Based Content Analysis. *Tijdschrift voor Ontwikkelingspsychologie, 22,* 61–74.

*Yuille, J. C. (1988a, June). *A simulation study of Criteria-Based Content Analysis*. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Yuille, J. C. (1988b). 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 and Psychiatry, 18,* 343–352.

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!