Let Me Inform You How to Tell a Convincing Story:

CBCA and Reality Monitoring Scores as a Function of Age, Coaching and Deception

Aldert Vrij[1][2]

Lucy Akehurst

Stavroula Soukara

Ray Bull

Running head: Coaching and Deception

[1] Correspondence should be addressed to: Aldert Vrij, University of Portsmouth, Psychology Department, King Henry Building, King Henry 1 Street, Portsmouth PO1 2DY, United Kingdom or via email: aldert.vrij@port.ac.uk

[2] This study was sponsored by a grant from the Economic and Social Research Council (R000222820)

Abstract

The first aim of this experiment was to examine whether being informed about a method of detecting deception called Criteria-Based Content Analysis (CBCA) would increase participants' CBCA scores when deceptive so that they might then be classified as truthful. The second aim was to investigate whether Reality Monitoring could be used as an alternative tool for verbal lie detection. The third aim was to examine whether participants' social skills (social anxiety, self monitoring and social adroitness) affected their CBCA scores. Participants (aged 6-8, 11-12, 14-15, and undergraduates) participated in a "rubbing the blackboard" event. In a subsequent interview they told the truth or lied about the event, after they were or were not taught some CBCA criteria. Truth tellers obtained higher CBCA scores than liars, and those who were informed about CBCA obtained higher scores than those who were not, except for the 6-8-year-olds. CBCA scores were also significantly correlated with social skills. Finally, Reality Monitoring was a useful alternative to CBCA for distinguishing between liars and truth tellers.

Let Me Inform You How to Tell a Convincing Story:
CBCA and Reality Monitoring Scores as a Function of Age, Coaching and Deception

To date, Criteria-Based Content Analysis (CBCA) -a systematic assessment of the credibility of written statements- is probably the most popular instrument to assess the veracity of written statements (Vrij, 2000).

CBCA is a systematic assessment of the credibility of written statements. Steller and Köhnken (1989) compiled a list of 19 criteria which had been used in such assessments. 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. This is known as the Undeutsch Hypothesis (Steller, 1989). The presence of each criterion strengthens the hypothesis that the account is based on genuine personal experience. Köhnken (1989, 1996, 1999, 2002) presented theoretical support for the Undeutsch hypothesis and proposed that both cognitive and motivational factors influence CBCA scores.

With regard to cognitive factors, it is assumed that, compared to those who fabricate a story, someone who actually experienced an event would be able to produce descriptions about this event which include more CBCA criteria, as some criteria (unstructured production, contextual embedding, reproduction of speech, unusual details, etc.) are believed to be very difficult for people to fabricate.

Other criteria are more likely to occur in truthful statements for motivational reasons. Truthful persons will not be as much concerned with impression management as will deceivers. Compared to truth tellers, deceivers would be more keen to try to construct a report which they believe will make a credible impression on others, and will leave out information which, in their view, will damage their image of being a truthful person (Köhnken, 1999). As a result, a truthful person's statement is more likely to contain information that is inconsistent with the beliefs/stereotypes that people have concerning truth

telling. The CBCA list includes several so-called "contrary-to stereotype" criteria (term adapted from Ruby & Brigham (1998)): "spontaneous corrections", "admitting lack of memory", "raising doubts about one's own testimony", etc..

CBCA was developed to evaluate statements from children. Many authors still describe CBCA as a technique solely developed to evaluate statements made by children in sexual offense trials (Honts, 1994; Horowitz, Lamb, Esplin, Boychuk, Krispin, & Reiter-Lavery, 1997). Others, however, advocate the additional use of the technique to evaluate the testimonies of adults (Köhnken, Schimossek, Aschermann, & Höfer, 1995; Porter & Yuille, 1996; Ruby & Brigham, 1997; Steller & Köhnken, 1989). These authors have pointed out that the underlying Undeutsch hypothesis is not restricted to children. The latter point of view has received some empirical support to date. Significantly higher CBCA scores for truth tellers than for liars have not only been found in studies with children witnesses (for example, Akehurst, Köhnken, & Höfer, 2001; Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997a, b; Lamers-Winkelman & Buffing, 1996; Tye, Amato, Honts, Kevitt, & Peters, 1999; Vrij, Akehurst, Soukara, & Bull, 2002; Winkel & Vrij, 1995) but also in studies with adult witnesses (for example, Akehurst et al., 2001; Höfer, Akehurst, & Metzger, 1996; Köhnken et al., 1995; Landry & Brigham, 1992; Porter & Yuille, 1996; Ruby & Brigham, 1998; Sporer, 1997; Vrij, Edward, & Bull, 2001a, b; Vrij et al., 2002; Zaparniuk, Yuille, & Taylor, 1995).

In the present study it was predicted that CBCA scores would be significantly higher for truth tellers than for liars in both adult statements and child statements (Hypothesis 1). It was further predicted that there would be a linear relationship between age and CBCA scores: the older the participants, the higher their CBCA score (Hypothesis 2). Theoretically, Hypothesis 2 could be explained in several ways. Cognitive abilities and command of language develop throughout childhood, making it gradually easier to provide detailed and rich accounts of what has been witnessed (Davies, 1991, 1994). Also, children are probably

less aware and/or less concerned with impression management than adults (Flavell, Botkin, Fry, Wright, & Jarvis, 1968; Vrij, 2002). Positive correlations between CBCA scores and age have been found in numerous studies before (see Vrij, in press, for an overview of these studies).

In principle, it is possible that if people were to learn which methods CBCA evaluators use to assess the credibility of their statements, they could try to "improve" their statements in order to obtain high CBCA scores which could then be assessed as truthful by CBCA judges. Previous experiments addressing this "coaching" issue have revealed that lying participants who were informed about CBCA were indeed capable of producing significantly higher CBCA scores than lying participants who were not informed (Vrij et al., 2002; Vrij, Kneller, & Mann, 2000). The present experiment also addressed the question of whether it is possible for people who have insight into the CBCA method to improve their CBCA score. The main difference between the previous experiments and this experiment was the age of the participants. Unlike the other experiments, we included a group of young participants (6-8-year-olds). We expected participants to benefit from the coaching and predicted that CBCA-coached participants would obtain higher CBCA scores than uncoached participants (Hypothesis 3).

Recently, Reality Monitoring has been used as an alternative method to examine verbal differences between responses believed to be true and false (Alonso-Quecuty, 1992, 1996; Alonso-Quecuty, Hernandez-Fernaud, & Campos, 1997; Höfer et al., 1996; Manzanero & Diges, 1996; Roberts, Lamb, Zale, & Randall, 1998; Sporer, 1997; Vrij, Akehurst, Soukara, & Bull, in press; Vrij, Edwards, Roberts, & Bull, 2000; Vrij et al., 2001a, b). The core of Reality Monitoring is the claim that memories of experienced events differ in quality from memories of imagined (e.g., fabricated) events. Memories of real experiences are obtained through perceptual processes and are therefore likely to contain, amongst others, perceptual information: details of smell, taste or touch, visual details and auditory details

(details of sound) and contextual information: spatial details (details about where the event took place, and details about how objects and people were situated in relation to each other, e.g., "He stood behind me") and temporal details (details about time order of the events, e.g., "First he switched on the video-recorder and then the TV", and details about duration of events). Accounts of imagined events are derived from an internal source and are therefore likely to contain cognitive operations, such as thoughts and reasonings ("I must have had my coat on, as it was very cold that night") (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981, 1998). One might argue that "experienced events" reflect truth telling whereas "imagined events" reflect deception. Therefore, differences between truth tellers and liars could be expected regarding Reality Monitoring criteria.

One of the benefits of the Reality Monitoring method, compared to CBCA, is that the method is relatively straightforward to use and less time consuming to apply (Sporer, 1997; Vrij et al., 2000, in press). Another benefit is that, unlike CBCA which consists of criteria solely related to truth telling, Reality Monitoring contains both truth telling criteria and a criterion indicative of deception (cognitive operations). However, Reality Monitoring needs to be more thoroughly tested. For example, previous studies often have failed to find the expected difference between liars and truth tellers regarding cognitive operations (Alonso-Quecuty, 1992, 1996; Höfer et al., 1996; Sporer, 1997; Vrij et al., 2000, 2001b). This might have been caused by the nature of the event. For example, in Vrij et al.'s (2000, 2001b) studies, participants were requested to give truthful or deceptive factual accounts of aspects of a film they had seen. This allows little room for cognitive operations. Truth tellers were asked to recall and liars were asked to fabricate what people in the film were doing, not what they, themselves, were thinking. Cognitive operations, however, are related to people's own thinking rather than recall of factual information about others. Therefore, the expected difference in cognitive operations between liars and truth tellers might occur when people are asked to describe their own activities during a certain period of time as this provides more

opportunity to include reports of cognitive operations. In support of this reasoning, Vrij et al. (in press) found that liars did include more cognitive operations in their accounts than truth tellers when they were asked to describe an event in which they were personally involved. In order to further strengthen the support for the cognitive operations hypothesis we sought to replicate this finding in the present study. It was hypothesised that truth tellers are likely to include more perceptual and contextual information in their statements than liars, and that liars are likely to include more cognitive operations in their statements than truth tellers (Hypothesis 4). For the same reason that we expected age differences in CBCA scores (i.e., cognitive abilities and command of language develop throughout childhood which makes it gradually easier to provide detailed accounts), we expected age differences in Reality Monitoring scores as well: The older the participants, the higher their Reality Monitoring score (Hypothesis 5).

Vrij et al. (2002) examined individual differences in CBCA scores. They argued that some people are more eloquent and verbally skilled than others, and that these differences may be related to social skills. They focused on three different social skills: social anxiety, social adroitness and self monitoring. People who are socially anxious feel discomfort in the presence of others (Buss, 1980) and their reports might therefore contain less quality (contain fewer CBCA criteria) than participants who feel more comfortable in social interactions. On the other hand, people who are socially adroit (Jackson, 1978) are experienced in verbally manipulating other people. As a result, their statements might contain more CBCA criteria than the statements of those who are less manipulative. Finally, people differ in the extent to which they naturally engage in impression management. Those who are high in self-monitoring (Snyder, 1987) are particularly concerned with making a favourable impression on others, and have the tendency to adjust their nonverbal and verbal behavior in order to create the desired effect on others. Therefore, it might be that high self monitors naturally produce higher CBCA scores. Vrij et al.'s (2002) findings supported these

assumptions. Their findings have potential implications. Statement Validity Assessment (SVA) experts typically look at alternative hypotheses to explain CBCA scores, such as cognitive limitations of the child, characteristics of the interview, and motivational factors. All these aspects are addressed in the Validity Checklist (Raskin & Esplin, 1991). Social skills of the interviewee, however, do not appear in the Validity Checklist and are therefore typically neglected. Vrij et al.'s (2002) findings suggest that social skills need to be incorporated in the Validity Checklist. However, the correlations found between CBCA scores and social skills were, although significant, not high (around r = .20). We therefore felt that a replication of these findings was desirable. Obviously, if the same findings were obtained, the plea to include social skills of the interviewee in the Validity Checklist would be strengthened. The relationship between social skills and Reality Monitoring scores have not been investigated to date. However, similar to CBCA scores, Reality Monitoring scores are probably related to eloquency, and we therefore expected the same relationships as for CBCA scores. In all, we predicted that social anxiety would be negatively correlated with CBCA scores and Reality Monitoring scores and that social adroitness and self monitoring would be positively correlated with CBCA and Reality Monitoring scores (Hypothesis 6).

Method

Participants

A total of 180 participants took part in the study, 92 (51%) males and 88 (49%) females. Their mean age was M=14.13 years, SD=5.56. There were four different age groups: 44 6 to 8-year-olds (13 (30%) males and 31 (70%) females, four were 6 years old, 30 were 7 years old, and 10 were 8 years old), 35 11 to 12-year-olds (21 (60%) males and 14 (40%) females, 16 were 11 years old and 19 were 12 years old), 44 14 to 15-year-olds (23 (52%) males and 21 (48%) females, 13 14-year-olds and 31 15-year-olds), and 57 undergraduate students (35 (61%) males and 22 (39%) females, their mean age was M=20.70 years, SD = 3.40).

Procedure

The experimental procedure was identical to Vrij et al.'s (2002, in press) procedures. The experiment took place at a Students' Union (for undergraduates) and at local schools (for children). Undergraduates were recruited under the guise of participating in an experiment about "telling a convincing story" with the possibility of earning £5. Children were asked by their teacher to go to see "a woman to play a game". (Prior to the study, parental informed consent was obtained). Participants took part individually. First, the undergraduates signed an informed consent form and filled out a questionnaire "about their personality" (which included self reports concerning social anxiety, social adroitness and self monitoring, see below). The 11-, 12-, 14- and 15-year-olds also filled out this questionnaire prior to the study. For the 6-8-year-olds, the questionnaire was sent to their parents (after the children had participated) who were asked to fill out the questionnaire on behalf of their children. After entering the experimental room, the female experimenter made fairly brief, polite conversation (exchanged names, what have you been doing?, etc.). From that moment events differed for the participants depending on which of the conditions they were in. Allocation to these conditions occurred at random.

Participants in the truthful condition (N = 86) played a game of "Connect 4" with the experimenter (all 180 participants in this study knew how to play this popular game). During the game, a person (we used different persons, see below) entered the room, said "Hello", and "Excuse me for interrupting", and also said "Ah! You are playing Connect 4, I'm hopeless at the game, I always lose!", walked to the blackboard and wiped some information (complicated math formulas) off the blackboard, and then left the room. After the game, the experimenter looked at the blackboard and then noticed that the information had been wiped off. She then asked the participant whether he/she saw someone wiping off the information. All participants in the truthful condition told the experimenter that the person who came in during the game wiped the information off the blackboard. Then, the experimenter gave the

following instructions: "Actually, I know that the information should not have been wiped off the blackboard, as it is needed for a lesson later on. In a minute you will be interviewed by another woman. Her task is to find out who wiped the information off the blackboard. Now, you know you did see who it was, so your task is to convince her that you did. All you need to do is be truthful about everything that happened while you were in this room. So say that we played a game of Connect 4 and that someone came in to wipe the blackboard. This is very important as, if you are successful in convincing her that you are telling the truth, we will give you (undergraduates - £5; 14-15-year-olds - £2; 11-12-year-olds - £2; 6-8-year-olds - a present). At the end of the interview, she will tell you whether she believes you or not. If she does believe that we played Connect 4 and that someone came in to wipe the blackboard, we will give you the money/present when you come out. If she doesn't believe you, you will not get any money/present at all. Do you understand?"[1]

Participants in the fabricating condition (N = 94) were told by the experimenter that she (the experimenter) earlier wiped some important information off the blackboard which was supposed to stay on there for a lesson later on. The participants were told that he/she would be interviewed by another woman whose task was to find out who wiped the information off the blackboard. The experimenter then asked the participants to pretend that it wasn't the experimenter who wiped the information off the blackboard. Instead the participants were asked to pretend that they (experimenter and participant) played Connect 4 and that someone else entered the room and wiped off the information during the game. Identical to the truthful witness condition, it was stressed that if the participant was successful in convincing the interviewer that it was someone else who wiped the information off the blackboard, he/she would receive £5/£2/present, and that he/she would receive nothing at all if the interviewer did not believe him/her. Again, participants were informed that the interviewer would tell the participants at the end of the interview whether or not she believed them.[2]

All participants were then randomly allocated to either the light coaching condition or the heavy coaching condition. Participants in the light coaching condition (N = 91) were told that it is important that the interviewer believes that the experimenter and participant played Connect 4 and that someone came in who wiped the information off the blackboard. It was further stressed that it is more likely that the interviewer would believe the participant if he/she told in lots of detail, what happened when he/she was in the room. For example, just saying: "Well, it was a bit boring really!" wouldn't convince the interviewer.

In addition to these light coaching instructions, participants in the heavy coaching conditions (N = 89) were given further information. The information differed for liars and truth tellers. Fabricating participants (except 6-8-year-olds) were given the following verbal instructions: "Here are some more things to think about when you are in the interview. Our research shows that these things help people to believe what you are saying: (a written version of these verbal instructions was put in front of the participants as well)

(1) try to include as many details as you can. For example, include some details about your "pretend" person. Was it a man or a lady? What was he/she wearing? (This is CBCA criterion 3).

(2) Describe "interactions", that is, describe how you think you would have reacted when we were playing the game or when the person interrupted us. For example, "I think (name of the experimenter) was a bit annoyed when I won the game" (CBCA criterion 5).

(3) Try to describe what people said to each other. For example, if someone had come into the room, what do you think he/she would have said to us? What do you think we would have said to one another when we were playing the game? (CBCA criterion 6).

(4) Pretend that something unusual or unexpected happened. For example, pretend that you dropped some of your discs for the game on the floor. (CBCA criterion 7).

(5) Describe how you felt when you were in the room. For example, if someone had interrupted our game, how do you think you would have felt? (CBCA criterion 12).

(6) Try to include how you think the person who wiped the blackboard was feeling."

Participants were then asked to think about the story they were going to tell in the subsequent interview. When they said they were ready to be interviewed, they were asked to practice the story with the experimenter first. The experimenter listened to the participants' story and checked whether the participant included all six criteria. If not, questions about these criteria were asked (i.e. "Could you think of something that the person who came in the room might have said to us?"). The interviewee was then encouraged to include this piece of information in their account as well. Finally, information about two more criteria (Criteria 14 and 15) were given:

(7) "The interviewer won't expect you to remember everything, because we can't always remember everything. If you don't remember something, tell the interviewer that you can't remember. Also, the interviewer will be more likely to believe you if you have forgotten some of the little things she asks you about. So it will be a good idea to tell her you have forgotten when she asks you. (CBCA criterion 15).

(8) Feel free to correct yourself if you think you have made a mistake, if you have contradicted yourself or something." (CBCA criterion 14).

The coaching for 6-8-year-olds was slightly different. They were only taught criteria 3, 5 and 6, and the training was conducted in a more interactive manner (i.e. "Pretend a woman came in: What was she wearing?" (child gives an answer); "What was the color of her hair?" (child gives an answer), etc.. The child was then encouraged to include the answers they had given in their upcoming interviews.

The heavy coaching condition for truthful participants was very similar to the heavy coaching condition just discussed. Obviously, they were not asked to pretend something but were simply asked to describe the actual event, i.e. were asked to describe what the person who wiped the blackboard was wearing, were asked to mention exactly what was said while the participant was in the room, etc..

After the coaching instructions were completed and after the participants had practiced their stories with the experimenter, the participant was brought to the (female) interviewer who was in another room. The interviewer was unaware of the condition the participant was in, except from the witness/suspect status of the participants (see endnote 2). (The interviewer needed to know the status of the participants as the instructions she had to give to suspects and witnesses differed). After building rapport with the participant, the interviewer gave participants in the witness condition the following instructions: "The reason I am interviewing you is that I have heard that someone wiped some important information off the blackboard in the room you were just in. I understand that you were in there just now with (name of experimenter). I need to know if you saw anything that will help me to find out who it was.[3] Also, remember that my colleagues need to know whether I think you are telling the truth or not. I would like you to tell me, in as much detail as possible, everything you can remember about what happened when you were in that room just now. Give me as much information as you can, even small details you do not think are very important".

After this free recall the interviewer told each participant that they had convinced her (the interviewer) that they told the truth and that they would receive the money/present. (For ethical reasons the interviewer was instructed to tell all children that she believed them, regardless of how convincing their stories were). With regard to the undergraduates, she was instructed not to give money to participants who gave very short statements. However, all undergraduates did put effort in giving extensive and credible statements, and all were paid £5. Outside, the participants received their money/present, were debriefed and thanked.

Questionnaire, CBCA Scoring, and Reality Monitoring scoring

Social anxiety was measured with Fenigstein, Scheier, and Buss' (1975) social anxiety scale (5 items). The questions for the 6-8-year-olds (which were answered by their parents) were written in a third person format. For example, the item "I get embarrassed very easily" (which was used for the 11-12-year-olds, 14-15-year-olds, and undergraduates) was

rephrased to: "The child gets embarrassed very easily"). Answers could be given on answer scales ranging from (1) untrue to (4) true (Cronbach's alphas were .86 (6-8-year-olds) and .77 (11-12-, 14-15-year-olds and undergraduates)). Social adroitness was measured with Jackson's (1978) social adroitness scale (20 items). Again, the child questions were written in a third person format (Cronbach's alphas were .77 (6-8-year-olds) and .76 (11-12-, 14-15-year-olds and undergraduates)). Self-monitoring was measured with Briggs, Cheeck, and Buss' (1980) self monitoring scale (21 items). Again, the child questions were written in a third person format (Cronbach's alphas were .70 (5-6-year-olds) and .63 (11-12-, 14-15-year-olds and undergraduates)). Pearson correlations revealed that self monitoring and social adroitness ($r(177) = .41$, $p < .01$) and self monitoring and social anxiety ($r(177) = -.45$, $p <$ 01) were significantly correlated, whereas social adroitness and social anxiety ($r(177) = -.12$, ns) were not.

Two independent raters received training in CBCA scoring. First, both raters read several major published papers about CBCA (Raskin & Esplin, 1991; Steller, 1989; Steller & Köhnken, 1989; Vrij, 2000; Vrij & Akehurst, 1998). Second, they were trained in CBCA scoring by a British CBCA expert. The expert explained each criterion under investigation in this study (see below) and gave examples of each criterion. Third, both the trainee raters and the expert rater evaluated one example transcript individually (from a different study). The three raters compared their results and feedback was given by the expert rater. Fourth, the trainees received more transcripts and were asked to rate these transcripts at home. In a follow up meeting, the results were evaluated and, again, feedback was given by the expert. After that meeting the expert felt that at that time the two raters had obtained sufficient rating skills and it was decided that the raters could commence their coding task for the present experiment. Coding was carried out individually by the two trained coders only (they coded the statements at home) and took place on the basis of the written transcripts of the interviews. The raters were blind to the hypotheses under investigation, to the staged event,

and to the experimental conditions the participants were allocated to (although they were aware that some transcripts would be truthful and some not). Some criteria ("accurately reported details misunderstood" (criterion 10); "pardoning the perpetrator" (criterion 18) and "details characteristic of the offense" (criterion 19)) were not scored, as they are specifically related to (sexual) crimes. "Self deprecations" (criterion 17) was initially to be scored but was never in fact present. This criterion was therefore disregarded, leaving a total of 15 CBCA criteria to be assessed. The coders scored the strength of presence of criteria 1 and 2 in each statement on 5-point Likert scales ((1) = absent, (5) is strongly present), and scored the frequency of occurrence of the other criteria in each statement. We then calculated a total CBCA score. This has been done before not only in experimental laboratory research (Tye et al., 1999; Vrij et al., 2000, 2001a, b, 2002, in press; Winkel & Vrij, 1995) but also in real life situations (Craig, Scheibe, Raskin, Kircher, & Dodd, 1999; Hershkowitz, Lamb, Sternberg, & Esplin, 1997; Lamb et al., 1997a, b; Parker & Brown, 2000). In order to create the CBCA scale the criteria were dichotomized. Dichotomizations for criteria 4 to 16 were based on the absence or presence of each of the criteria in the interview. A score of 0 was assigned when the criterion was absent, and a score of 1 when the criterion was present. For criteria 1 and 2, a 0 was assigned when the criterion obtained a "1" rating on the 5-point Likert scale, and a 1 was assigned when the criterion obtained a score of 2, 3, 4, or 5 on the 5-point Likert scale. For dichotomization of criterion 3, median splits were used.[4] Total CBCA scores were calculated for both coders. The correlation between these two CBCA scores was moderate but acceptable (.66). In the present analyses we used as total CBCA score the average score of the two coders. The score could range from 0 to 15.[5]

Two other raters received training in Reality Monitoring (RM) scoring. A British RM expert (different from the CBCA expert) provided the raters with a detailed description of how the criteria should be scored, including some case examples. Then, both the trainee raters and the expert evaluated some example transcripts individually (from a different

study). The three raters compared their results and feedback was given by the expert. At this stage the expert and the two raters felt that the raters were capable of scoring the transcripts without any further instructions. This is in agreement with Sporer (1997) who also found that it is much easier to teach (and to learn) RM scoring than CBCA scoring. With regard to the present study, coding was carried out individually by the two trained raters (they coded the statements at home) and involved the written transcripts of the interviews. The raters were blind to the hypotheses under investigation, to the staged event, and to the experimental condition (although they were aware that some transcripts would be truthful and some not). The two raters scored per interview the frequency of occurrence of visual details (e.g., "I walked in to the room" contains three visual details), auditory details (e.g., "She said to sit down" contains one sound detail), temporal details (e.g., "We started playing" is one temporal detail), spatial details (e.g., "And then the pieces fell on to the floor" contains one spatial detail) and cognitive operations (e.g., "Because she was quite clever, she won the game" contains one cognitive operation; so do "I presume that the two people knew each other" and "She was quite tall for a girl"). Following previous examples (Vrij et al., 2000, 2001b, in press) a total Reality Monitoring score was calculated. In order to create the Reality Monitoring scale the visual, auditory, spatial and temporal variables were dichotomized (see also Vrij et al., 2000, 2001b, in press). Dichotomizations for auditory, temporal and spatial details were based on the absence or presence of each of the criteria in the interview. A score of 0 was assigned when the criterion was absent, and a score of 1 when the criterion was present. Cognitive operations were dichotomized as well, with a score of '1' given if no cognitive operations were present and a score of '0' when cognitive operations were present. For dichotomization of visual details, median splits were used. For reasons explained in the CBCA scoring paragraph (see above), different median splits were used for 6-8-year-olds and the other participants. Total Reality Monitoring scores were calculated for both coders. The correlation between these two Reality Monitoring scores was satisfactory (.71). In the present

analyses we used as total Reality Monitoring score the average score of the two coders.[6] The Reality Monitoring scale contained five criteria (visual details, auditory details, spatial details, temporal details and cognitive operations) and the total-score could range from 0 to 5.[7]

Results

CBCA Scores and Reality Monitoring Scores as a Function of Truth-status, Age, and Coaching

In order to test Hypotheses 1 to 5, a MANOVA was conducted with Truth-status (truth vs lie), Age (6-8, 11-12, 14-15, undergraduates) and Coaching (light vs heavy) as factors and CBCA score, and Reality Monitoring score as dependent variables. At a multivariate level, the analysis revealed main effects for all three factors, Truth-status, $F(2, 163) = 23.62$, $p < .01$, Age, $F(6, 324) = 8.84$, $p < .01$, and Coaching, $F(2, 163) = 6.26$, $p < .01$. None of the interaction effects were significant.[8]

At a univariate level, the Truth-status factor revealed significant findings for both dependent variables. In support of Hypotheses 1 and 4, truth-tellers obtained higher CBCA scores ($M = 6.73$, $SD = 1.5$) than liars ($M = 6.19$, $SD = 1.7$), $F(1, 164) = 6.24$, $p < .01$, and truth-tellers obtained higher Reality Monitoring scores ($M = 3.77$, $SD = .8$) than liars ($M = 2.79$, $SD = 1.0$), $F(1, 164) = 47.49$, $p < .01$.

Regarding the Age factor, univariate tests revealed significant findings for CBCA scores, $F(3, 164) = 17.18$, $p < .01$. The linear trend, predicted in Hypothesis 2, emerged (see Table 1). The older the participants, the higher their CBCA scores. The CBCA scores for 14-15-year-olds and undergraduates were significantly higher than the CBCA scores for 6-8-year-olds; 11-12-year-olds obtained significantly lower CBCA scores than undergraduates. The CBCA scores for 6-8- and 11-12-year-olds did not differ significantly from each other, neither did the CBCA scores for 14-15-year-olds and undergraduates. Univariate tests revealed a significant effect for Reality Monitoring score as well, $F(3, 164) = 3.35$, $p < .05$.

However, Tukey HSD tests did not reveal differences between the four age groups. One could argue that this is caused by the inclusion of cognitive operations in the Reality Monitoring score, as there is no reason to assume that younger children would include more cognitive operations in their statements than older children or undergraduates (this is necessary to obtain a higher Reality Monitoring score). We therefore excluded cognitive operations from the total Reality Monitoring score for this analysis (the new total Reality Monitoring score consisted of four dependent variables: visual, auditory, spatial and temporal details). We conducted Tukey HSD tests on this newly constructed Reality Monitoring score and the findings are shown in Table 1. The pattern which emerged is very similar to the CBCA pattern just described, and supports Hypothesis 5.

Regarding the Coaching factor, univariate tests revealed one significant effect. As was predicted in Hypothesis 3, heavily coached participants obtained higher CBCA scores (M = 6.79, SD = 1.7) than lightly coached participants (M = 6.12, SD = 1.5), $F(1, 164)$ = 10.89, $p < .01$.

Of theoretical interest are the Truth-status X Coaching and Coaching X Age interaction effects. Although the Truth-status X Coaching effect regarding CBCA scores was not significant, $F(1, 164) = .82$, ns, an interesting pattern emerged. The difference between truth tellers and liars was not significant for the heavily coached participants, $F(1, 87) = .82$, ns, but was significant for the lightly coached participants, $F(1, 89) = 5.41$, $p < .05$. In the lightly coached condition truth tellers obtained a higher CBCA score (M = 6.50, SD = 1.4) than liars (M = 5.80, SD = 1.4).

The Truth-status X Coaching interaction effect was significant for Reality Monitoring scores, $F(1, 164) = 5.02$, $p < .05$. In the heavily coached condition, truth tellers obtained a higher Reality Monitoring score (M = 3.60, SD = .9) than liars (M = 2.95, SD = .9, $F(1, 87)$ = 11.38, $p < .02$. The same pattern emerged in the lightly coached condition, $F(1, 89) = 46.86$, $p < .01$, with truth tellers obtaining higher Reality Monitoring scores (M = 3.95, SD = .7)

than liars (M = 2.65, SD = 1.0). The differences between liars and truth tellers were larger in the lightly coached than in the heavily coached condition.

Although the Coaching X Age interaction effect regarding CBCA scores and Reality Monitoring scores was not significant, F(3, 164) = 1.35, ns and F(3, 164) = .31, ns respectively, examining the results per age group is relevant, particularly for the CBCA scores. The results are presented in Table 2. The results for the 6-8-year-olds showed that coaching was not successful. CBCA scores for heavily coached and lightly coached participants were almost identical and did not differ significantly from each other. For the three remaining groups, however, the CBCA scores were significantly higher for heavily coached participants than for lightly coached participants.

CBCA Scores, Reality Monitoring Scores and Social Skills

In order to test Hypothesis 6, Pearson correlations were carried out between CBCA scores and Reality Monitoring scores on the one hand and social anxiety, social adroitness, and self monitoring on the other hand. The results are shown in Table 3.

As was predicted in Hypothesis 6, CBCA scores were positively correlated with social adroitness and self monitoring scores and negatively correlated with social anxiety scores. All three correlations were significant. Regarding Reality Monitoring scores one significant finding emerged. Reality Monitoring scores were negatively correlated with social anxiety. A distinction between liars and truth tellers (see Table 3) revealed that the correlation patterns, just described, particularly occurred when participants were lying.[9]

CBCA Scores and Reality Monitoring scores: Their Discriminative Power

In order to determine the usefulness of the detection techniques in classifying truth tellers and liars, stepwise discriminant analyses utilizing the Wilks' Lambda method were conducted. With this technique the variables remaining in the final analysis are those which contribute to maximizing the correct assignment of the cases to the objective truth status. In the analyses, the objective truth status was the classifying variable and the CBCA score, and

Reality Monitoring score were the dependent variables. In order to examine whether CBCA classifications would benefit from a 'lie criterion', the original, non-dichotomized, cognitive operations variable was introduced as a dependent variable as well. Because coaching had an impact on the dependent variables, analyses were carried out for the lightly coached participants only. The results are given in Table 4.

As can be seen in Table 4, 60% of the cases could be correctly classified on the basis of the CBCA scores alone, whereas 74% could be correctly classified on the basis of the Reality Monitoring alone. The analysis which included CBCA scores plus cognitive operations revealed that 74% of the cases could be correctly classified on the basis of these two variables (both cognitive operations, Wilks' Lambda = .83, and CBCA score, Wilks' Lambda = .71, contributed to the discriminant function). Finally, the analysis in which CBCA and Reality Monitoring scores were included showed that only Reality Monitoring contributed significantly to the discriminant function.[10]

Discussion

The experiment revealed that truth tellers obtained higher CBCA scores and Reality Monitoring scores than liars. There is now a substantive number of studies showing that CBCA scores and Reality Monitoring scores do discriminate between liars and truth tellers (see Vrij (2000) for reviews). The finding regarding cognitive operations (presented in endnote 8) is more exceptional. We have now demonstrated for the second time that, conforming to predictions, liars, compared to truth tellers, include more cognitive operations in their account when they describe a life-event (rather than recalling a film they have seen). The discriminant analyses revealed that liars and truth tellers were more accurately classified on the basis of their Reality Monitoring scores than on the basis of their CBCA scores. These findings support those who claim that Reality Monitoring could be used as an alternative tool of verbal lie detection. As mentioned in the Introduction, one of the advantages of Reality Monitoring is that the method also contains a lie criterion (cognitive operations). The CBCA

technique (which only scores for truth) might become more balanced and might result in higher accuracy when lie telling criteria (criteria which indicate deception) are included. The present findings support this assumption. A discriminant analysis which included the CBCA scores and cognitive operations as dependent variables resulted in a higher hit rate (74%) than the discriminant analysis with CBCA scores alone (60%). Also Vrij et al. (in press) found that accuracy improved when cognitive operations was included as an additional variable alongside CBCA. The analyses further revealed that more accurate classifications of liars and truth tellers were made if the cognitive operations variable was removed from the total Reality Monitoring scale and was introduced as a separate variable alongside this new Reality Monitoring scale which did not contain cognitive operations (see endnote 10). This is likely due to the fact that the cognitive operations variable is conceptually different from the other criteria: its presence does not indicate truth telling (as is the case with the other criteria), but lying. **Treating the two sets of variables as distinctive sets of variables could therefore be recommended.**

A second aim of the experiment was to examine whether participants could improve their CBCA scores when they were given insight into the CBCA criteria. The findings revealed that coaching indeed improved participants' CBCA scores, however, not for the youngest group of participants (6-8-year-olds). This will please those who support the use of CBCA assessments regarding young children being used as evidence in criminal court. However, our findings do not suggest that teaching young children to improve their CBCA scores is impossible. The "training" was short (it lasted less then ten minutes) and perhaps young participants would benefit from a more extensive training session. However, we do believe that our findings suggest that training young children to improve their CBCA scores is difficult.

Our final aim was to investigate to what extent CBCA scores are influenced by participants' social skills. The same findings emerged as in our previous study (Vrij et al.,

2002). CBCA scores were positively correlated with social adroitness and self monitoring and negatively correlated with social anxiety. These findings also emerged when we controlled for the effects of the coaching manipulation and age of the participants. On the basis of these two studies, we therefore recommend SVA evaluators to take people's social skills into account when they evaluate CBCA scores. This relationship between social skills and CBCA scores was clearer in lying participants than in those who were telling the truth. The same pattern emerged in Vrij et al.'s (2002) study. Perhaps this has to do with the difficulty of the task. It seems plausible to suggest that the task was more difficult for liars than for truth tellers. Perhaps a task needs to be verbally challenging in order for these social skills focusing on verbal skills to discriminate.

The inter-rater agreements between the two coders regarding the individual CBCA criteria were (i) sometimes low and (ii) mostly lower than their inter-rater agreement regarding the total CBCA scale. These patterns are consistent with findings in previous CBCA research (see Vrij, in press, for a discussion about these issues). It implies that, at least on reliability grounds, experts involved in real-life cases should use the total CBCA scale rather than relying on individual CBCA criteria. Although we are not aware of CBCA experts who merely focus on individual criteria in real life cases, we cannot rule out that experts who do not officially use the CBCA method but who are familiar with its content would do so.

Finally, five issues merit attention. First, different age groups received different rewards, therefore effects of age were confounded with type of motivation. We varied the reward so as to avoid such a confound. We believe that giving all participants the same reward would have created a confound because the same reward (for example £5) would be perceived as substantially higher by young children than by undergraduates. However, it is difficult to determine what would be comparable rewards for different age groups. With hindsight, we could have asked participants this. Second, the coaching manipulation for the 6-8-year-olds differed somewhat from the coaching manipulation of the other groups of

participants. Although this created a confound in experimental terms, we had no other choice. In a previous study (Vrij et al., 2002), we exposed young children to the heavy coaching condition we used for older participants in this study but had to cancel this condition as some children clearly did not understand the instructions. Third, we did not include a non-coaching condition in our experimental design. We did so for ecological validity reasons. In experimental studies participants are probably less motivated to make a convincing impression than in real life situations. We were therefore afraid that not providing any guidance to our participants might have resulted in poor quality statements. Obviously, comparing a heavy coaching condition with a no coaching condition would have made significant findings more likely, but even in the present study, where we compared heavy coaching with light coaching, we found the predicted coaching effects on CBCA scores. Fourth, following our previous study (Vrij et al., 2000), the parents rather than the children themselves completed the social skills questionnaires in the 6-8-year-old age group. Asking such young children to complete the questionnaires themselves was impossible for obvious reasons. We have no reasons to believe that the results were affected by this possible shortcoming. For example, the predicted relationships between social skills and CBCA scores occurred even when the impact of 'age' was partialled out. Also, the parents were blind to the hypotheses under investigation. Fifth, some sceptics will argue that the outcomes of this study say little about SVA assessments in the field, as we only focused on CBCA and not on the whole lengthy SVA procedure. They will say that CBCA should not be assessed independent of the other components of the procedure. We disagree with those sceptics for several reasons. First, although CBCA is only part of the SVA procedure it is the core part of that procedure. Clearly, the outcome of SVA assessments should be received with high scepticism if the core of the procedure (CBCA) is found not to discriminate between truth tellers and fabricators. Second, a qualitative review of the application of statement analyses in Swedish courts (one of the countries where SVA originated) showed that experts rely

heavily on the CBCA outcomes (Gumpert & Lindblad, 2001). That is, although they sometimes acknowledge that several external factors (typically those factors covered in the Validity Checklist) might have affected a CBCA outcome, they never came to the conclusion that these factors could solely explain the CBCA scores and therefore high-quality statements (which result in high scores) were generally interpreted as referring to self-experienced events.

References

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

Alonso-Quecuty, M. L. (1992). Deception detection and Reality Monitoring: A new answer to an old question? In F. Lösel, D. Bender, & T. Bliesener (Eds.), Psychology and law: International perspectives (pp. 328-332). Berlin: Walter de Gruyter.

Alonso-Quecuty, M. L. (1996). Detecting fact from fallacy in child and adult witness accounts. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), Psychology, law, and criminal justice: International developments in research and practice (pp. 74-80). Berlin: Walter de Gruyter.

Alonso-Quecuty, M. L., Hernandez-Fernaud, E., & Campos, L. (1997). Child witnesses: Lying about something heard. In S. Redondo, V. Garrido, J. Perez, & R. Barbaret (Eds.), Advances is psychology and law (pp. 129-135). Berlin: Walter de Gruyter.

Buss, A. H. (1980). Self-consciousness and social anxiety. San Francisco: Freeman.

Briggs, S. R., Cheek, J. M., Buss, A. H. (1980). An analysis of the Self-Monitoring Scale. Journal of Personality and Social Psychology, 38, 679-686.

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.

Davies, G. M. (1991). Research on children's testimony: Implications for interviewing practice. In C. R. Hollin and K. Howells (Eds.), Clinical approaches to sex offenders and their victims (pp. 93-115). Chichester, Great Britain: Wiley & Sons.

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

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. Journal of Consulting and Clinical Psychology, 43,

522-527.

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

Gumpert, C. H., & Lindblad, F. (2001). Expert testimony on child sexual abuse: A qualitative study of the Swedish approach to statement analysis. Expert Evidence, 7, 279-314.

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, Sienna, 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., 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.

Jackson, D. N. (1978). Interpreter's guide to the Jackson Personality Inventory. In P. McReynolds (Ed.), Advances in psychological assessment, volume 4 (pp. 56-102). San-Francisco: Jossey-Bass.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. Psychological Bulletin, 114, 3-29.

Johnson, M. K., & Raye, C. L. (1981). Reality Monitoring. Psychological Review, 88, 67-85.

Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. Trends in Cognitive Sciences, 2, 137-145.

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, NJ: Springer-Verlag.

Köhnken, G. (1996). Social psychology and the law. In G. R. Semin, & K. Fiedler (Eds.), Applied Social Psychology (pp. 257-282). London, Great Britain: Sage Publications.

Köhnken, G. (1999, July). Statement Validity Assessment. Paper presented at the pre-conference programme of applied courses 'Assessing credibility' organised 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, United Kingdom: 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.

Lamb, M. E., Sternberg, K. J., Esplin, P. W. Hershkowitz, I., Orbach, Y., & Hovav, M. (1997a). Criterion-based content analysis: A field validation study. Child Abuse and Neglect, 21, 255-264.

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

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 Publications.

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.

Manzanero, A. L., & Diges, M. (1996). Effects of preparation on internal and external memories. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), Psychology, law, and criminal justice: International developments in research and practice (pp. 56-63). Berlin, Germany: Walter de Gruyter.

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.

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.

Raskin, D. C., & P. W. Esplin (1991). Statement Validity Assessment: Interview procedures and content analysis of children's statements of sexual abuse. Behavioral Assessment, 13, 265-291.

Roberts, K. P., Lamb, M. E., Zale, J. L., & Randall, D. W. (1998, March). Qualitative differences in children's accounts of confirmed and unconfirmed incidents of sexual abuse. Paper presented at the biennial meeting of the American Psychology-Law Society, Redondo Beach.

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.

Snyder, M. (1987). Public appearances/ private realities: The psychology of self-monitoring. New York: Freeman.

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 (1989). Credibility assessment (pp. 135-154). Deventer, the Netherlands: Kluwer.

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.

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, 92-109.12

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Ed.), Handbuch der Psychologie Vol. 11: Forensische Psychologie (pp. 26-181). Göttingen, Germany: Hogrefe.

Vrij, A. (2000). Detecting lies and deceit: The psychology of lying and the implications for professional practice. Chichester, Great Britain: John Wiley & sons.

Vrij, A. (2002). Deception in children: A literature review and implications for children's testimony. In H. Westcott, G. Davies, & R. Bull (Eds.), Children's testimony (pp. 175-195). Chichester, Great Britain: Wiley & Sons.

Vrij, A. (in press).Criteria-Based Content Analysis: A qualitative review of the first 37 studies. Psychology, Public Policy, and Law.

Vrij, A., & Akehurst, L. (1998). Verbal communication and credibility: Statement Validity Assessment. In A. Memon, A. Vrij, & R. Bull, Psychology and law: Truthfulness, accuracy and credibility (pp. 3-26). Maidenhead, Great Britain: McGraw-Hill.

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. (in press). Detecting deceit via analyses of verbal and nonverbal behaviour in children and adults. Human Communication Research.

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., 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.

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

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.

Table 1

CBCA scores and Reality Monitoring scores as a Function of Age (N = 180)

	Age								
	6-8		11-12		14-15		undergraduates		
	m	sd	m	sd	m	sd	m	sd	
CBCA	5.30[a]	1.1	6.23[ab]	1.6	6.75[bc]	1.6	7.25[c]	1.5	
RM[3]	3.09[b]	.9	2.22[a]	1.1	2.56[a]	.9	3.05[b]	.7	

[3] The total Reality Monitoring score consisted of four variables: visual, auditory, temporal and spatial details.

Table 2

CBCA scores as a function of Coaching and Age

	Coaching			
CBCA scores	light		heavy	
	m	sd	m	sd
6-8	5.29	1.0	5.31	1.1 F(1, 42) = .00
11-12	5.63	1.3	6.94	1.7 F(1, 33) = 6.49*
14-15	6.36	1.4	7.13	1.7 F(1, 42) = 2.77*[4]
undergraduates	6.86	1.5	7.64	1.4 F(1, 55) = 3.87*

* $p < .05$

[4] one-tailed test

Table 3

Pearson Correlations between CBCA score, Reality Monitoring score, and Social Anxiety-, Social Adroitness-, and Self Monitoring Scores

	Social Anxiety	Social Adroitness	Self Monitoring
Total group (N = 177)			
CBCA	-.25**	.23**	.24**
RM	-.23**	-.03	.03
Liars (N = 92)			
CBCA	-.39**	.29**	.26*
RM	-.28**	.13	.09
Truth tellers (N = 85)			
CBCA	-.02	.21[5]	.26*
RM	-.05	-.11	.08

* p < .05 ** p < .01

[5] p = .056 (two-tailed)

Table 4.

Discriminant Analyses with Criteria-Based Content Analysis, Reality Monitoring and Cognitive Operations: Lightly coached participants only (N = 91).

	hit rates						
Detection technique	lie	truth	total	Eigenvalue	Lambda	df	X2
CBCA	50%	60%	.06	.94	1	5.41	69%
RM	61%	88%	74%	.52	.65	1	37.43
CBCA + co	74%	74%	74%	.41	.71	2	30.09
CBCA + RM	61%	88%	74%	.52	.65	1	37.43

* p < .05, ** p < .01

1. It might seem suspicious from the participant's perspective that the experimenter would know that an interview was about to ensue in order to ascertain the truth of the blackboard mishap. We had prepared an answer which the experimenter would give in case a participant asked a question about this. However, none of the participants in this study raised this issue. In other words, we have no evidence that the participants were suspicious.

   Telling the participants at the end of the interview whether or not the interviewer believed the participants reflects police practice: such statements sometimes occur at the end of suspect and witness interviews.

2. The truthful and deceptive conditions described are the witness conditions. Similar to Vrij et al. (2002) we also introduced suspect conditions for the three oldest groups of participants. The truthful suspect condition was similar to the truthful witness condition, but this time participants were accused of having wiped the information off the blackboard themselves. In the deceptive suspect condition, participants had wiped off the blackboard themselves and were accused of having done so in the subsequent interview. In addition to the promised gift if they succeeded in telling a convincing story, truthful and deceptive suspects also faced a punishment if they were not convincing, they were be asked to write down in detail what has happened. This suspect/witness manipulation did not yield a significant effect and is disregarded in this article. See Vrij et al. (2002) for a theoretical rationale for this manipulation.

3. The interviewer gave participants in the suspect condition the following instructions: "The reason I am interviewing you is that I have heard that YOU wiped some important information off the blackboard in the room you were just in. I understand that you were in there just now with (name of the experimenter). I need to know whether it was you or not!" For the remaining part, the suspect interviews and instructions were identical to the witness interviews and instructions.

4. Separate median splits were used for (i) 6-8-year-olds and (ii) for the remaining participants for the following reason: An ANOVA with Age Group as factor and details as dependent variable (scores of 2 coders combined) revealed a significant effect, $F(3, 176) = 18.28$, $p < .01$. Tukey HSD tests revealed that 6-8-year olds included significantly fewer details in their statements ($M = 14.75$, $SD = 5.8$) than any of the other three groups (10-11-year-olds: $M = 21.69$, $SD = 8.3$; 14-15-year-olds: $M = 24.80$, $SD = 8.2$; undergraduates: $M = 26.99$, $SD = 10.6$). These three groups did not differ significantly from each other. A median split for the whole group would therefore imply that almost all 6-8-year-olds would be allocated to the "low score" group and many other participants to the "high score" group (as the median split would be relatively low).

5. Inter-rater agreement were also calculated for the individual criteria on the original, non-dichotomized, scores. Several (Pearson) correlations were high: reproduction of conversation ($r = .83$), quantity of details ($r = .76$), and admitting lack of memory ($r = .74$); others were moderate: raising doubt about one's own testimony ($r = .58$), description of interactions ($r = .52$), contextual embeddings ($r = .51$), accounts of own mental state ($r = .45$), superfluous details ($r = .44$), accounts of other's mental state ($r = .42$); others were rather low but significant: unexpected complications ($r = .35$), spontaneous corrections ($r = .35$), and logical structure ($r = .34$). The remaining correlations were not significant: unstructured production ($r = .20$), related external associations ($r = .11$), and unusual details ($r = .09$). The latter two criteria were hardly present, and appeared in less than 8% of the statements. Previously (Vrij et al., 2002), we left out the criteria with low reliability before computing a total CBCA score. However, this does not reflect real-life practice where CBCA experts typically rely on all the criteria they assess. We therefore included all 15 criteria in the total CBCA score, and the correlation between the total CBCA scores of both judges (.66) justified this. We return to this issue in the Discussion.

6. Inter-rater agreements were also calculated for the five individual criteria. All Pearson correlations were satisfactory: auditory details ($r = .96$), temporal details ($r = .87$), visual details ($r = .76$), spatial details ($r = .68$) and cognitive operations ($r = .61$).

7. All truthful participants participated in more or less the same staged event. It might therefore be that the truths told by these participants bore certain similarities. These similarities could be picked up by the CBCA and Reality Monitoring raters after a few trials of coding, and this "knowledge" might have affected their codings. We do not think that this actually happened.

Although there were similarities in the staged event, there were also differences which were purposefully introduced by us in order to prevent this happening. For example, different people were used to come into the room and wipe the blackboard. Also, the same person wore different clothes at different times. As a result, the descriptions of the "actor" differed considerably even in the truthful condition. (Because variations of the same staged event were introduced, we videotaped all participants while they were in the room with the experimenter, and checked the veracity of their stories afterwards by comparing their statements with what actually had happened. We did not come across any commissions, purposefully distorting the truth, in the truthful reports). Additionally, not all truthful participants gave a complete account, with some participants describing some features and other participants describing totally different features.

8. For reasons pointed out in the Introduction, we were also interested in the results for the cognitive operations variable separately. An ANOVA was conducted with Truth-status, Age and Coaching as factors and cognitive operations (the original, not dichotomized scores) as dependent variable. The analysis revealed main effects for Truth-status, $F(1, 164) = 38.56$, $p < 01$), and Age, $F(3, 164) = 7.18$, $p < .01$), and a Truth-Status X Age interaction effect, $F(3, 164) = 3.88$, $p < .05$. The interaction revealed that liars included significantly more cognitive operations in their accounts than truth tellers in all age groups except in the youngest age group (6-8-year-olds: $M = .41$, $SD = .9$ vs $M = .10$, $SD = .3$, $F(1, 42) = 2.37$, ns; 11-12-year-olds: $M = .56$, $SD = .5$ vs $M = .22$, $SD = .4$, $F(1, 33) = 4.99$, $p < .05$; 14-15-year-olds: $M = 1.17$, $SD = 1.1$ vs $M = .40$. $SD = .6$, $F(1, 42) = 8.12$, $p < .01$; undergraduates: $M = 1.55$, $SD = 1.3$ vs $M = .24$, $SD = .4$, $F(1, 55) = 26.09$, $p < .01$). Tukey HSD tests regarding the Age main effect revealed that 6-8-year-olds included fewer cognitive operations in their accounts ($M = .26$, $SD = .7$) than the other age groups which did not differ significantly from each other (11-12-year-olds: $M = .39$, $SD = .5$; 14-15-year-olds: $M = .82$, $SD = 1.0$; undergraduates: $M = .93$, $SD = 1.2$).

9. Partial correlations controlling for (i) the coaching manipulation and (ii) age of the participant revealed the same significant effects as the effects presented in Table 3, suggesting that age and coaching have no moderated impact on the relationships between social skills and CBCA and Reality Monitoring scores.

10. In subsequent analyses, we separated cognitive operations from the Reality Monitoring scale (leaving the Reality Monitoring scale with four variables: visual, auditory, temporal and spatial details) and included the new Reality Monitoring scale and cognitive operations (original, non-dichotomized scale) as two separate dependent variables in the discriminant analysis. This analysis yielded higher accuracy rates than the accuracy rates presented in Table 4 (lie: 76%, truth: 88%, total 81%) and both Reality Monitoring scores (Wilks' Lambda = .82 and cognitive operations, Wilks' lambda = 56, contributed significantly to the discriminant function). A discriminant analysis with total CBCA score, total (new) Reality Monitoring score and cognitive operations as dependent variables gave high accuracy rates too (lie: 76%, truth: 93%, total: 84%). All three variables contributed to the discriminant function (Reality Monitoring score, Wilks' Lambda = .82, cognitive operations, Wilks' Lambda = .56, CBCA score, Wilks' Lambda = .55).

Analyses per age group are available from the first author.