The Effects of Training Professional Groups and Lay Persons to use Criteria-Based Content Analysis to Detect Deception

LUCY AKEHURST¹*, RAY BULL¹, ALDERT VRIJ¹ and GÜNTER KÖHNKEN²

¹University of Portsmouth, UK
²University of Kiel, Germany

SUMMARY
This experiment was designed to assess, for the first time, the effects of training police officers, social workers and students in Criteria-Based Content Analysis (CBCA) in an attempt to increase lie detection accuracy. A within-subjects design was implemented. Participants rated the truthfulness of a maximum of four statements before training in CBCA and rated the truthfulness of a different set of four statements after training. The raters were only exposed to the written transcripts of the communicators. Two thirds of the statements utilized were truthful and one third were based on fabrications. Before training, there were no significant differences in detection accuracy between the police officers (66% accuracy), the social workers (72% accuracy) and the students (56% accuracy). After training, the social workers were 77% accurate and significantly more accurate than the police officers (55%) and the students (61%). However, none of the three groups of raters significantly improved their lie detection accuracy after training, in fact, the police officers performed significantly poorer. Overall, police officers were significantly more confident than social workers and lay persons regardless of accuracy. Further, participants were most confident when labelling a statement truthful regardless of whether or not this was the correct decision. Copyright © 2004 John Wiley & Sons, Ltd.

Research on the detection of deception via verbal and nonverbal behavioural cues has shown that the ability of untrained raters to discriminate between true and deceptive statements is generally not very good. Vrij (2000, in press) reported that the average accuracy of lie detection, in studies with a chance expectation of 50%, varies within a range of 45% to 60%. However, even 60% accuracy is insufficient for the judgement of eyewitness statements in real court cases since, at best, four out of ten judgements would be wrong (Köhnken, 1987a).

Judgements of credibility based on the contents of a statement have, in general, produced higher accuracy rates (than nonverbal behaviour) ranging from 65% to 90% (Vrij, 2000, in press). One method of assessing the truthfulness of a statement, based on its content, was developed in German forensic psychology by Undeutsch (1967, 1984) and

*Correspondence to: Dr. Lucy Akehurst, Department of Psychology, University of Portsmouth, King Henry I Street, Portsmouth PO1 2DY, UK. E-mail: lucy.akehurst@port.ac.uk

Copyright © 2004 John Wiley & Sons, Ltd.
Arntzen (1983). Based on their work, Köhnken and Steller (1988) and Steller and Köhnken (1989) compiled a list of criteria and described a procedure for evaluating the veracity of a statement which has led to the development of Statement Validity Assessment (SVA). The major component of the SVA technique, Criteria-Based Content Analysis (CBCA) focuses on specific content characteristics which, if present in a statement, support the hypothesis that the account is based on genuine personal experience (i.e. that it is truthful).

TRAINING TO DETECT DECEPTION USING CBCA

As Köhnken (1987a) stated, ‘an interesting question, of considerable practical importance, deals with the trainability of lie detection skills’ (p. 4). The poor accuracy rates of untrained lay people should not be a concern if it were possible to enhance detection accuracy by training them to detect deception. The same applies for professionals (e.g. police officers and social workers) who need to assess the credibility of others as part of their job. There are very mixed findings with regard to the trainability of lie detection skills based on nonverbal cues with little evidence to suggest that training actually works (see e.g. DePaulo, Lassiter, & Stone, 1982; DeTurk, 1991; DeTurk, Feeley, & Roman, 1997; Kassin & Fong, 1999; Köhnken, 1987a; Porter, Woodworth, & Birt, 2000).

Köhnken (1987a) was the first to investigate whether knowledge of the CBCA criteria would improve police officers’ ability to detect deception. He divided his trainees into four groups. Group one was instructed to attend to the spoken content of the statements they observed (using a selection of the CBCA criteria). Group two was instructed to pay attention to certain speech behaviour (behaviours found by Köhnken to discriminate between truths and lies), group three was instructed to attend to aspects of nonverbal behaviour (again, empirically based), and group four was given no specific instructions.

The mean overall accuracy rate was 45% and no effect of instructions was found. Köhnken suggested that one of the reasons for these disappointing findings may have been that the statements classed as ‘deceptive’ did in fact contain some elements of the truth. This is likely to be common in real life police statements and is seldom addressed in the deception literature. Köhnken also suggested that training to detect deception could result in cognitive overload resulting in reduced accuracy rates.

Apart from the study by Köhnken (1987a), there has been very little, methodologically sound, research concerning the effects of training in CBCA with regard to credibility assessment and those that have considered this issue have used lay persons as their trainees. Steller, Wellershau, and Wolf (1988), in an attempt to recreate the emotional and experiential qualities of real life sexual abuse, asked 88 children to relate truthful and fabricated stories concerning episodes such as an attack by an animal, medical treatment, and getting lost in a shopping mall. In 1989, Steller used these statements in a subsequent study. Three raters trained in CBCA and three untrained raters (undergraduate students) evaluated the verbatim, written statements with regard to their truthfulness. CBCA training, although it only lasted 90 minutes, produced 78% correct classifications for truthful statements and 62% correct classifications for deceptive statements (average 70%). The untrained raters were able to correctly classify 68% of truthful statements and 47% of fabricated statements (average 57.5%). The group differences were statistically significant.
Landry and Brigham (1992) carried out a study to assess the effectiveness of training in CBCA on the ability to detect deception in adults. Training merely involved a 45-minute session based on material supplied by two ‘experts in the field’. Participants were either shown videorecordings of communicators (six truthful adults and six fabricating adults) or were shown the written transcripts of the same communicators. Trained raters who saw videotapes performed significantly better than any of the other groups but still only achieved an accuracy rate of 58%. The untrained raters who used videotapes achieved 51% accuracy, the trained raters who used transcripts achieved 53% accuracy, and the untrained raters who used transcripts achieved 42% accuracy. Likely reasons for these particularly poor accuracy rates include the rather inadequate training programme and the small sample of communicator statements.

OLD DOGS AND NEW TRICKS

Putting the training of lay persons to detect deception using CBCA aside, there have been no published research studies addressing the success (or otherwise) of training professional groups to detect lies using CBCA. The first aim of this study was to examine the differences between the accuracy of ‘experienced’ lie detectors working with children, (i.e. police officers and social workers) and students in detecting deception via the written transcripts of children before and after training in CBCA. A one-day training package was used. It was hypothesized that police officers and social workers might achieve higher accuracy rates (i.e. correct decisions about the truthfulness of statements) than would students with no equivalent everyday need to judge the credibility of children (Hypothesis 1).

It is reasonable to assume that training in CBCA, a technique that when used by experts has been shown to elicit accuracy rates of up to 90% (Akehurst, Köhnken, & Höfer, 2001), may feasibly increase police officers’ and social workers’ credibility assessment accuracy.

However such a training programme should be approached with caution. Memon, Holley, Milne, Köhnken, and Bull (1994) looked at the effects of relatively brief training of experienced police officers in the cognitive interview, a technique that had been repeatedly shown, when used by similarly trained students acting as interviewers in laboratory settings, to increase correct detail reported by cooperative witnesses (see Milne & Bull, 1999 for a review). The police officers did not increase the amount of correct detail reported by interviewees. Memon et al. suggested that the officers had difficulty (i) applying the new technique and (ii) giving up their old methods of interviewing.

As training to detect deception has historically involved a great deal of ‘myth dissolution’ (e.g. fidgeting and eye contact have not been found to be associated with lying) it is perhaps too much to expect that a new technique introduced over a relatively short period of time will overcome some of the ingrained prejudices (Vrij, 2000) that hamper some professionals in their day to day judgements of credibility.

Further, Köhnken (1987b) suggested that CBCA is a technique that should be carefully trained by experienced psychologists and, in general, that this should take around three weeks. Köhnken further recognized that to train this complicated technique in a short period of time may simply confuse participants and actually hinder their lie detection accuracy. However, since 1987, research has been published which has found that brief training in CBCA can improve accuracy rates beyond those found when nonverbal cues
are used (for a review see Vrij, in press). It is not viable to spend up to three weeks training police officers and social workers to use CBCA, a shorter training package would be more readily accepted by these professionals. With the findings of Steller (1989) and Landry and Brigham (1992) in mind, it was hypothesized that all three groups of participants would improve in their lie detection accuracy after training in CBCA (Hypothesis 2). However, we were somewhat tentative in this expectation due to the valid issues raised by Köhnken (1987b).

CONFIDENCE AND JUDGEMENTS OF CREDIBILITY

Vrij (2000) and Meissner and Kassin (2002) found that police officers were significantly more confident in their judgements of credibility than lay persons. However, their accuracy rates did not mirror these inflated ratings of confidence. In line with these findings it was hypothesized that the police officers and social workers in this study would be more confident in their evaluations than the lay persons (Hypothesis 3). It was further hypothesized that, for all participants, ratings of confidence would be higher after training in CBCA than before (Hypothesis 4). It is likely that when participants feel ‘armed’ with a new technique the fact that they can focus on specified cues will make them more confident than when they are simply basing their judgements on an overall gut feeling with no specific strategy. In sum, the authors were interested in how profession and training influenced confidence levels and did not investigate, for this paper, changes in the confidence-accuracy relationship.

In the present study, police officers, social workers and students judged the credibility of different samples of children’s statements before and after training in CBCA thus utilizing a within subjects design. To examine whether the ratings of the content criteria (after training) actually significantly discriminated between truthful and fabricated statements—it was hypothesized that the content criteria would be rated as appearing significantly more often in the truthful statements than in the fabricated statements (Hypothesis 5) as suggested by the originators of the technique (Arntzen, 1983; Köhnken & Steller, 1988; Steller & Köhnken, 1989; Undeutsch, 1984).

METHOD

Participants

There was a total of 58 participants in this experiment. Twenty-six police officers from three different constabularies took part. Twenty were on child protection and domestic violence training courses and six were taking part, with the social workers, in a joint training course on issues surrounding child witnesses.1 All the police officers had served at least five years. Their mean age was 36.46 years (SD = 6.09 years), there were 9 males and 17 females. Of these police officers 13 had read literature concerning detecting deception, however none were familiar with CBCA as a means for judging the credibility of statements.

1The experimenter ‘took over’ the courses in question for one day and all attendees on the courses took part in the investigation.
Fourteen social workers from one county also took part. They were all taking part in a joint training course (as mentioned above). All the social workers had been in their profession for at least 5 years. Their mean age was 44.42 years ($SD = 6.20$ years), there were 2 males and 12 females. Of these social workers 3 had read literature concerning detecting deception, however none were familiar with CBCA.

Eighteen undergraduate students also took part. The were recruited from a psychology department participant pool at a British university and received course credit for participation. Their mean age was 24.17 years ($SD = 7.70$ years), there were 6 males and 12 females. Of these students, none had read literature concerning detecting deception and none were familiar with CBCA.

**Stimulus material**

The stimulus materials used to assess the ability of participants to detect deception were children’s statements, either truthful or based on fabrication, regarding interaction with a professional photographer (originally collected for Akehurst et al., 2001). The children were 49 7–11 year old children ($M = 9.2$ years, $SD = 0.47$ years), 23 males and 26 females. The children were randomly assigned to one of three experimental conditions; (i) truthful: actual involvement, (ii) truthful: watched video and (iii) fabrication. One third of the children were assigned to the ‘truthful: actual involvement’ condition and they took part in a photography session and had three photos taken, each child being photographed separately. The sessions were exactly the same for each child and lasted approximately five minutes. All the photography sessions were videotaped. These children were informed that they would be interviewed the next day about what had happened.

The children assigned to the ‘truthful: watched video’ condition watched a video-recording of a member of the same age group having his/her photograph taken. All the children watched the same session in small groups. These children were also informed that they would be interviewed the next day about what they had seen.

The children assigned to the ‘fabricating’ condition were given information about the photography session verbally by a confederate of the experimenter. They were given a set of verbal cues as follows, to help them understand what had happened in the photography sessions. For example, ‘the photographer was a young man’ and ‘he took three photographs’ (see Akehurst, Köhnken, & Höfer, 2001).

These children in the ‘fabricating’ condition were also informed that they would be interviewed the next day and that their task would be to give a description of the photography session so that the interviewers would not know that they had not in fact had their photograph taken. The experimenter asked the children to imagine having their photograph taken in as much detail as possible and to fill in any appropriate details that were not mentioned in the verbal description.

No time constraints were imposed and each interview was audio and videorecorded. The audiorecordings were then used to obtain verbatim written transcripts/statements of each interview. The mean word length for the three types of account were calculated. For ‘truthful: actual involvement’ $M = 421.53$ words ($SD = 214.52$ words), for ‘truthful: watched video’ $M = 458.86$ words ($SD = 152.46$ words) and for the ‘fabricated’ statements $M = 340.61$ words ($SD = 286.98$ words). A univariate ANOVA was performed on the data with type of statement as independent variable and number of words as dependent
variable. No significant difference was found with regard to the length of truthful and fabricated statements \( (F(2, 48) = 1.13, \text{n.s.}) \).

**Procedure**

The training session was conducted with groups ranging from 8 to 20 people. All the participants rated a randomly selected sample of four statements before training in CBCA and then a different selection of four statements after training. This number was decided upon so as to provide enough attempts to make the statistical analysis sufficiently powerful whilst ensuring that the experimental sessions were not too long.

The statements (i.e. truthful or not) were randomly assigned to packages of four statements. Therefore it was possible for participants to rate any combination of truthful and fabricated statements \( (N = 4) \) ranging from all truthful to all fabricated. Participants were informed of this random allocation to prevent them making assumptions regarding the possible combinations of statements they received. However, it was ensured by the investigators that equal numbers of truthful and fabricated statements were rated before and after training for each of the groups, i.e. police officers, social workers and students.

Initially all the participants were told that the present study was designed to investigate how well people can tell the difference between truthful and fabricated children’s statements.

**Before training instructions**

‘You will shortly be given a package containing four written statements and four rating forms. Please rate each statement with regard to its truthfulness. The statements are all derived from interviews with children aged 7–11 years old. The age of the child and his/her gender is marked on each statement. The statements concern the visit of a professional photographer to the children’s school. In all the statements the children are either talking about having their photograph taken or watching a video of another child having his/her photograph taken. Some of the statements are truthful i.e. the children really did have their photograph taken or they watched a video of another child having his/her photograph taken and are therefore based on real-life recollections. However some of the statements are fabricated, that is some children were coached to pretend they had their photograph taken or they watched a video of a photography session, and were asked to fool the interviewer into believing they were telling the truth. Please read each statement in the package given to you carefully. You can read each statement as often as necessary. There are no time constraints, please do not confer with others or swap statements with another person. Please rate the truthfulness of each statement on the forms provided.’

The rating forms asked the participants to give two ratings for each statement. Firstly they were simply asked, Do you think this statement is truthful (yes or no)? Secondly they were asked to rate how confident they were about their decisions on 7-point Likert scales ranging from zero (not at all confident) to six (completely confident).

**Training**

Having rated their first four statements the participants were informed that these initial evaluations were to act as a baseline concerning their present ability to detect deception in
children’s statements. They were informed that they would be trained in a relatively new technique to the UK that may help them in the future to assess the truthfulness of statements. All the participants were then given a four hour training session in CBCA. The criterion attribution of perpetrator’s mental state was amended to attribution of photographer’s mental state to fit with the stimulus event used in the present study.

The trainer began by introducing the SVA technique. The technique’s origins and present status were discussed and each phase of the approach was outlined. The majority of the training was, however, devoted to the content criteria that comprise the CBCA phase of the overall technique. Each participant was given a booklet that summarized the 13 content criteria used in the present study. The trainer went over each criterion, explaining it in depth and giving examples. It was emphasized, on many occasions, that the presence of these criteria suggested truthfulness of a statement. Three rules proposed by Steller (1989) were also presented, firstly mere repetitions in different passages within a statement do not increase the rating of a criterion, secondly one passage in a given statement can fulfil more than one criterion, and thirdly only contents that are in some way related to the event (i.e. the photography session) should be considered.

The training ended with a practice session in which the participants were given two statements from a ten-year old, completely unrelated to the present study (initially the trainees were not told that these practice statements were truthful). The trainees were asked to read through their statements and to rate the presence of each of the 13 criteria that had been discussed on 5-point Likert scales ranging from zero (absent) to four (strongly present) and then to give an overall assessment of the truthfulness of the statements. It was suggested that trainees carry out the task individually at first but then discuss their evaluations with others who had rated the same statements. The trainer (who had herself undergone a rigorous, three week training course concerning CBCA, conducted by an expert in the field who regularly gives expert testimony regarding the credibility of child witnesses using SVA) had previously rated each of the practice statements, highlighting pieces of text that related to the criteria, and had given a rating for the presence of each criterion. The trainer, with each group, compared her ratings with those of the trainees and discussions with each of the small groups were carried out until agreement of ratings for each of the criteria was reached to within 1 point on the scales. The trainees were informed, after they had completed their ratings, that the statements were truthful.

After training instructions

‘You will shortly be given a second package containing four written statements and four rating forms. Again, please rate these statements with regard to their truthfulness. These statements will be different from those you rated earlier and have been randomly allocated, therefore whatever the combination of truthful and fabricated statements you considered you received previously does not relate in any way to this selection. The statements do, however, relate to the same event, that is the photography session and are from the same age range of children i.e. 7 to 11 year olds. Also, some of the statements are fabricated and some are truthful as was explained earlier. Please read each statement in the package given

---

2Six of the original CBCA criteria were not utilized, namely, accurately reported details misunderstood, related external associations, accounts of subjective mental state (should be crime-related), self-deprecation, pardoning the perpetrator, and details characteristic of the offence. These were not used because they are specific to reports of sexual abuse.
to you carefully. You can read each statement as often as necessary. There are no time constraints, please do not confer with others or swap statements with another person. This time you are required to use the content criteria we have discussed today to inform your decisions regarding the truthfulness of the statements. Remember the more criteria that are present in a statement the more likely that statement is to be truthful. Feel free to refer back to the booklets that describe the criteria. Please rate the truthfulness of each statement on the forms provided.

The rating forms in the ‘after training’ packages required the participants to give ratings for the presence of each of the 13 content criteria that had been discussed and practised earlier in the session. The criteria were rated on 5-point Likert scales ranging from zero (absent) to four (strongly present) as was the case for the practice statements. The participants were also asked to give the same two types of ratings that they made before training (i.e. they were asked whether the statements were truthful (yes or no), and they were asked to rate how confident they were about their decisions).

**RESULTS**

Due to time constraints not every participant rated four statements before training and four statements after training. However, the minimum any participant rated was three statements before training and three statements after training. A total of 423 observations were made (from 48 statements), 214 before training (139 truthful and 75 fabricated) and 209 after training (135 truthful and 74 fabricated).

**Accuracy scores**

Overall percentage accuracy scores were computed using the yes/no responses, regarding the truthfulness of each statement, that each participant made for each statement they rated.

To test Hypothesis 1 (i.e. that police officers and social workers would achieve higher accuracy rates than students) and Hypothesis 2 (i.e. that detection accuracy would be higher after training in CBCA) a repeated measures analysis of covariance (ANCOVA) was carried out with profession of the participants as a between subjects independent factor (police officers vs social workers vs students) and when they rated the statements (i.e. before vs after training) as a within subjects independent factor. Gender and age of trainees were included in the analysis as covariates to control for any differences between the three groups (police officers, social workers and students). The overall percentage accuracy scores were the dependent variables. There was no significant main effect for when statements were rated (i.e. before or after training) ($F(1, 49) = 1.12$, n.s.). There was no significant main effect for the profession of participants ($F(2, 49) = 1.54$, n.s.). Finally there was no interaction effect between profession of participant and when statements were rated.

---

3To give an example of the quality of statements given; sections of text relating to descriptions of interactions included, ‘he moved my arm as he couldn’t see it properly’, and ‘I helped him with the curtains because it was too light’.

4The investigators checked that there was not an accidental discrepancy in the numbers of true and false statements rated before and after training by the three groups. It was found that every statement (total = 48) was rated four to six times before training and four to six times after training.
were rated \(F(2, 49) = 1.85, \text{n.s.}\). Therefore, there was no support for Hypotheses 1 and 2. When each profession, i.e. police officers, social workers and students, was considered separately there was a significant difference in the accuracy of police officers before and after training. However, the police officers were more accurate before training (68% accuracy) than they were after training (53% accuracy) \(F(1, 25) = 4.97, p < 0.05\). Students did achieve increased accuracy after training (57% accuracy before vs 62% accuracy after), but this effect was not significant (see Table 1 for all means).

Confidence ratings

To test Hypothesis 3 (i.e. that professionals would be more confident than lay persons), and Hypothesis 4 (i.e. that participants would increase in confidence after training in CBCA) a repeated measures ANCOVA was carried out with when the statements were rated (i.e. before vs after training) as a within subjects independent factor and profession of the participants (police officers vs social workers vs lay people), accuracy of the rating (accurate vs inaccurate) and status of the statement (truth vs lie) as between subjects independent factors. Gender and age of trainees were included in the analysis as covariates. The mean confidence scores (measured on Likert scales, 0–6) were the dependent variables. There was no main effect for when statements were rated (i.e before training \(M = 4.36\) vs after training \(M = 4.51\) \(F(1, 385) = 0.96, \text{n.s.}\)) therefore, Hypothesis 4 was not supported. However, there was a significant main effect on confidence with regard to the profession of participants \(F(2, 385) = 5.80, p < 0.05\). Tukey HSD tests were carried out to see where the significant differences occurred. It was found that police officers were significantly more confident (\(M = 4.81\)) than social workers (\(M = 4.00\)), \(p = 0.002\), and than lay people (\(M = 4.38\)), \(p = 0.05\). There was no significant difference in the confidence scores for lay people and social workers. This partially supports Hypothesis 3.

One significant interaction effect was found concerning the accuracy of judgements and whether they were for truthful or fabricated statements \(F(2, 385) = 18.19, p < 0.01\). Further analyses were carried out to see where the significant differences occurred. It was found that when truthful statements were rated, participants were significantly more confident in their accurate judgements (\(M = 4.92\)) compared to their inaccurate judgements (\(M = 3.90\)), \(F(1, 273) = 27.34, p < 0.01\), however when fabricated statements were rated participants were more confident in their inaccurate judgements (\(M = 4.91\)) compared to their accurate judgements (\(M = 4.25\)) \(F(1, 148) = 7.25, p < 0.01\). Interestingly, the interaction effect illustrates that participants were confident when they correctly classified truthful statements but were also confident when they erroneously classified fabricated statements as truthful. That is, they were most confident when they said ‘true’

---

**Table 1. Percentage accuracy scores for the three groups before and after training**

<table>
<thead>
<tr>
<th>Profession</th>
<th>Before training</th>
<th>After training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Police officers</td>
<td>68</td>
<td>53</td>
</tr>
<tr>
<td>Social workers</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Students</td>
<td>57</td>
<td>62</td>
</tr>
</tbody>
</table>
regardless of whether this was an accurate decision. (see Table 2 for all the mean confidence ratings).

**Ratings of the CBCA criteria**

To test Hypothesis 5 (i.e. that the ratings of the content criteria would significantly discriminate between truthful and fabricated statements, appearing more often in the truthful statements) a multivariate analysis of variance (MANOVA) was conducted. The type of statement (truthful: actual involvement vs truthful: watched video vs fabricated) was the independent variable and the ratings of the criteria were the dependent variables. At a multivariate level, there was a significant difference between the ratings of criteria for the different types of statement, $F(26, 376) = 2.26, p < 0.005$. At a univariate level, four criteria showed a significant effect of type of statement; contextual embedding, $F(2, 200) = 3.51, p < 0.05$, quantity of detail, $F(2, 200) = 8.41, p < 0.001$, reproduction of conversation, $F(2, 200) = 8.28, p < 0.001$ and description of interactions, $F(2, 200) = 3.39, p < 0.05$. Tukey HSD tests were carried out to ascertain where the significant differences lay. For contextual embedding, the criterion appeared significantly more often in the true (actual involvement) condition than in the fabrication condition. There were no significant differences between the other combinations (i.e. true (video) vs fabrication and true (actual involvement) vs true (watched video)). For quantity of detail, the criterion appeared significantly more often in the true (watched video) condition than in the fabrication condition. There were no significant differences between the other combinations. For reproduction of conversation the criterion appeared significantly more often in the true (actual involvement) and true (watched video) conditions compared to the fabrication condition. There was no significant difference between the two truthful conditions and finally, for description of interactions, the criterion appeared significantly more often in the true (actual experience) condition than the fabrication condition. There were no significant differences between the other combinations. In sum, the differences were in the expected direction, that is, the criteria appeared more often in the truthful statements compared to the fabricated statements thus partially supporting Hypothesis 6 (see Table 3 for means).

**DISCUSSION**

Hypothesis 1 of the present study stated that police officers and social workers would achieve higher accuracy rates (i.e. correct classifications regarding the truthfulness or
otherwise of statements) than lay persons. Before training, there were no significant differences between the percentage accuracy of each of the groups. It should be noted here that the social workers were unusually accurate in their decisions regarding the truthfulness of statements even before training. Their 75% level of accuracy is much higher than the 50% level that can be expected by chance alone and reasons for their initial high accuracy are discussed later. However, after training there were significant differences between the percentage accuracy of the professions. The social workers were still the most accurate (75% accuracy) followed by the students (62% accuracy) and the police officers performed poorest after training (53% accuracy) Thus there was little support for Hypothesis 1.

After the training sessions the experimenter was able to discuss the CBCA technique and the format of the training sessions with the participants. Two points may explain the Hypothesis 1 findings. Firstly, it became apparent that the social workers were a lot more knowledgeable and more experienced in evaluating children and especially in dealing with allegations of abuse. Six of the police officers were taking part in a joint training programme with social workers to improve procedures for dealing with crimes that involve children as witnesses or victims, thus they had experience of such cases. However, twenty of the police officers (the majority) were taking part in an initial training course for child protection and domestic violence. They were therefore relatively inexperienced with such cases and were only just beginning to gain some understanding of this area. Similarly, the students had no experience of working with children as witnesses or victims. It is probable that to benefit fully from training in CBCA, trainees must have a good knowledge of the development of children (e.g. familiarity with norms relating to social and cognitive development and language acquisition) and should be experienced in recognizing the ways in which children disclose abuse and subsequently talk about it.

Secondly, the majority of the police officers admitted to feeling somewhat ‘bamboozled’ by the training session, they said that they felt they needed to take the material away and digest it slowly before they would be confident to use the technique. This anecdotal evidence supports Köhnken’s (1987a) comments regarding the possibility of

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Truthful: actual involvement mean (N = 77)</th>
<th>Truthful: watched video mean (N = 58)</th>
<th>Fabricated mean (N = 74)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical consistency</td>
<td>3.86 (0.90)</td>
<td>3.79 (1.15)</td>
<td>3.51 (1.18)</td>
</tr>
<tr>
<td>Unstructured production</td>
<td>2.82 (1.28)</td>
<td>2.84 (1.06)</td>
<td>2.46 (1.11)</td>
</tr>
<tr>
<td>Quantity of detail</td>
<td>3.44 (1.21)</td>
<td>3.83 (1.09)</td>
<td>3.03 (1.19)</td>
</tr>
<tr>
<td>Contextual embedding</td>
<td>2.57 (1.31)</td>
<td>2.31 (1.40)</td>
<td>2.04 (1.16)</td>
</tr>
<tr>
<td>Description of interactions</td>
<td>3.26 (1.35)</td>
<td>3.40 (1.15)</td>
<td>2.89 (1.21)</td>
</tr>
<tr>
<td>Reproduction of conversation</td>
<td>2.91 (1.46)</td>
<td>3.19 (1.38)</td>
<td>2.24 (1.36)</td>
</tr>
<tr>
<td>Unexpected complications</td>
<td>1.87 (1.22)</td>
<td>1.83 (1.14)</td>
<td>1.72 (1.10)</td>
</tr>
<tr>
<td>Unusual details</td>
<td>2.43 (1.32)</td>
<td>2.57 (1.23)</td>
<td>2.39 (1.27)</td>
</tr>
<tr>
<td>Superfluous details</td>
<td>2.25 (1.40)</td>
<td>2.33 (1.32)</td>
<td>1.93 (1.26)</td>
</tr>
<tr>
<td>Attribution of photographer’s mental state</td>
<td>1.68 (0.98)</td>
<td>1.48 (0.84)</td>
<td>1.65 (0.90)</td>
</tr>
<tr>
<td>Spontaneous corrections</td>
<td>1.65 (1.01)</td>
<td>1.57 (0.92)</td>
<td>1.87 (1.08)</td>
</tr>
<tr>
<td>Admitting lack of memory</td>
<td>3.10 (1.26)</td>
<td>3.05 (1.32)</td>
<td>3.49 (1.23)</td>
</tr>
<tr>
<td>Raising doubts</td>
<td>1.84 (1.20)</td>
<td>1.91 (1.11)</td>
<td>2.03 (1.24)</td>
</tr>
</tbody>
</table>

*Standard deviations in parentheses.
‘cognitive overload’ when training to detect deception. The police officers may have performed poorer than the students because they were not used to learning about such a complicated technique in such a short period of time. Students, who are more used to being ‘taught’ new ideas and theories may have found it easier to assimilate the information and put it into practice. Thus, the police officers may have performed poorest after training because they were ‘out of practice’ at being taught and the majority still had little experience of dealing with children.

In relation to the point made above, that the police officers would have preferred to take the material away to digest before they used the technique, this problem could have been overcome by allowing all the participants to take their second batch of statements home to rate in their own time. There are however potential problems inherent with this suggestion. Firstly, the experimenter would not have had any guarantee that the participants would send back the completed rating forms once they had evaluated the statements. Secondly, participants may have discussed their statements with one another and finally, there would have been no uniformity with regard to how long each participant took to rate their statements. Therefore, although the police officers felt rushed during the training sessions there was no other way to ensure that the rating of the second batch of statements was controlled and uniform across all participants.

Hypothesis 2 stated that observers would increase their detection accuracy after training in CBCA. Overall, there was no significant difference between the accuracy scores before and after training in CBCA. When each profession, i.e. police officers, social workers and students, was considered separately there was a difference in the accuracy of police officers before and after training, however, they performed poorer after training. These findings do not support Hypothesis 2.

There are two reasons that may account for the finding that training did not increase detection accuracy. Firstly, the low overall detection accuracy might indicate that in fact CBCA is not a successful tool for judging the truthfulness of statements. However, previous studies (Akehurst et al., 2001; Horowitz et al., 1997; Köhnken, Schimossek, Aschermann, & Höfer, 1995; Lamb et al., 1997; Porter & Yuille, 1996; Vrij, Akehurst, Soukara, & Bull, 2002; Vrij, Edward, Roberts, & Bull, 2000) do not support this reasoning. All these studies have found evidence to support the fact that CBCA can be used successfully to discriminate truth from fabrication. It is interesting to note that the ratings of the presence of the 13 criteria used in the present study did successfully discriminate between the truthful and fabricated statements (see later).

It would appear that even after rating the statements for the presence of the criteria, the trainees were not able to ‘translate’ their ratings into accurate overall judgements of the truthfulness of the statements. It would therefore seem essential that, in future, CBCA training packages pay more attention to decision rules. That is, trainees should be given more guidance regarding how to relate their ratings of criteria to their overall decisions concerning the truthfulness of the statements (a technique shown to be successful by Vrij, Evans, Akehurst, & Mann, 2004).

5It should be noted that the differences between the criteria ratings for truthful and fabricated statements for this study were small. Had there genuinely been small differences between the true and false statements, this may have had a negative impact on accuracy rates. However, a previous piece of research (Akehurst et al., 2001) using the same statements, found larger differences when experts were conducting the CBCA analysis thus negating this possibility.
Secondly, it is perhaps not surprising that the CBCA training was not successful given the limited time and materials that were utilized and the difficult task that the observers faced (as was reflected in the overall poor accuracy rate). As mentioned in the introduction, Köhnken (1987b) stated that CBCA was a complicated technique and it was likely that significant increases in accuracy might only be expected after comprehensive training (around three weeks). It is possible that more in-depth training and more exposure to the criteria would increase the percentages of correct classifications by a significant amount. However, it may be unrealistic to expect professional groups to resource such extensive training.

A reason for the decrease in accuracy for police officers after training in CBCA may be connected to the point made previously and those made by Memon et al. (1994). From a subjective standpoint the police officers appeared to be the group of trainees that found the technique the hardest to understand. Before training they were using methods that they had used previously to detect deception in written statements and were correct 66% of the time. A reason for their reduction in accuracy after training may be that the task was further complicated for them by learning about the CBCA criteria. The ‘short’ course in CBCA may have simply confused them and acted in a detrimental manner with regard to detection accuracy.

Vrij (2000) and Meissner and Kassin (2002) found that professionals who had to detect deceit as part of their jobs were more confident in their responses regardless of accuracy. This relationship was tested in Hypothesis 3 and partial support was found for it. When the confidence ratings were considered for each of the professions separately it was found that, overall, police officers were significantly more confident than social workers and students but there was no significant difference in the confidence scores for students and social workers. The social workers were not confident about their inaccurate responses (before and after training) possibly because they had a good knowledge of the development of children and were experienced in recognizing the ways in which children disclose abuse and subsequently talk about it. Thus, with this greater experience compared to the police officers and students, they were less sure of the judgements they got wrong.

Hypothesis 4 stated that observers would become more confident after training in CBCA regardless of their accuracy. Although, not statistically significant, it was noted that there was a trend for increased confidence after training. However, this increased confidence was not associated with an increase in accuracy, as discussed previously.

DePaulo et al. (1997) found, in a meta-analysis, that participants were more confident when rating truthful compared to fabricated statements, however we did not find this. What we found was a significant interaction between accuracy and type of statement (truthful or fabricated). Participants were significantly more confident when they were accurate about true statements (i.e. they said ‘true’ and were correct) and when they were inaccurate about fabricated statements (i.e. they said ‘true’ but were incorrect).

Hypothesis 5 stated that the ratings of the content criteria would significantly discriminate between truthful and fabricated statements, appearing more often in the truthful statements. There was a significant difference between the ratings of criteria for the different types of statement. Overall, the criteria appeared significantly more often in the truthful statements compared to the fabricated statements, thus supporting Hypothesis 5.

It should be noted that generalization from the present study to on-the-job lie detection abilities of police officers and social workers should be approached with caution. The low detection accuracies may indicate that the judgement procedures and materials, particularly the statements used for this study, may not have been sufficiently similar to those of
real witness statements. All the participants commented on how difficult it seemed to them to assess the truthfulness of the children simply by assessing their written statement. It could be argued that in real life, numerous other cues (i.e. nonverbal cues and paralinguistic cues) are available and more detailed case information is available to help an evaluator make a decision about the truthfulness of a witness or victim. Very little research has been conducted concerning nonverbal cues to deception in children (see Vrij, 2002). However, if the literature concerning nonverbal cues to deception in adults proves to be comparable, it seems that nonverbal cues can often be very misleading and unhelpful when making judgements of credibility (see Vrij, 2000 for an overview).

Overall, these findings have shown that, after a relatively brief training in CBCA, neither professionals nor students improved their ability to detect deception when asked explicitly to do so. The participants were fully aware throughout the experiment that their task was to eventually rate the credibility of each child. However, the results of the MANOVA showed that the ratings that the participants made with regard to the presence of the CBCA criteria did significantly discriminate truths from lies. It was the translation of these ratings into a final overriding assessment (true or false) that caused problems.

Previous research concerning implicit methods of detecting deception has shown that participants are better able to distinguish between the truth and fabrication when they are not asked explicitly to make the decision truth or lie (DePaulo, Anderson, & Cooper, 1999; Vrij, Edward, & Bull, 2001; Vrij et al., 2004). For example, Vrij et al. (2004) firstly asked participants to judge the frequency of occurrence of a number of verbal and nonverbal behaviours (including CBCA criteria) occurring during a statement and no mention was made of credibility at this stage. Subsequently, the participants were most accurate in their credibility judgements when they were asked to make a decision regarding truthfulness based purely on the frequency ratings they had made.

REFERENCES


