

Running head: Using an example statement with truth tellers and liars

This is the peer reviewed version of the following article: Bogaard, G., Meijer, E. H. and Vrij, A. (2014), Using an Example Statement Increases Information but Does Not Increase Accuracy of CBCA, RM, and SCAN. *J. Investig. Psych. Offender Profil.*, 11: 151–163. doi: 10.1002/jip.1409, which has been published in final form at <http://onlinelibrary.wiley.com/doi/10.1002/jip.1409/full> This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

**Using an example statement increases information but does not increase
accuracy of CBCA, RM and SCAN**

Glynis Bogaard^{*}, Ewout H. Meijer

Department of Clinical Psychological Science, Maastricht University

Aldert Vrij²

Department of Psychology, University of Portsmouth

Acknowledgement. This research was funded by a grant from the Ministry of the Interior and Kingdom Relations of the Netherlands.

*Requests for reprints should be addressed to Glynis Bogaard, Dept. of Clinical Psychological Science, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands (e-mail: Glynis.bogaard@maastrichtuniversity.nl).

Abstract

Verbal credibility assessment methods are frequently used in the criminal justice system to investigate the truthfulness of statements. Three of these methods are Criteria Based Content Analysis (CBCA), Reality Monitoring (RM) and Scientific Content Analysis (SCAN). The aim of this study is twofold. First, we investigated the diagnostic accuracy of CBCA, RM, and especially SCAN. Secondly, we tested whether giving the interviewee an example of a detailed statement can enhance the diagnostic accuracy of these verbal credibility methods. To test the latter, two groups of participants were requested to write down one true and one fabricated statement about a negative event. Prior to this request, one group received a detailed example statement, whereas the other group received no additional information. Results showed that CBCA and RM scores differed between true and fabricated statements, whereas SCAN scores did not. Giving a detailed example statement did not lead to better discrimination between truth tellers and liars for any of the methods, but did lead to the participants producing significantly longer statements. The implications of these findings are discussed.

Keywords

Criteria Based Content Analysis, Reality Monitoring, Scientific Content Analysis, verbal lie detection, deception detection, credibility assessment methods

Using an example statement increases information but does not increase accuracy of CBCA, RM and SCAN

Research has shown that in general, people are not good at detecting lies, and that even trained police officers generally perform only just above chance level (Aamodt & Custer, 2006; Bond & DePaulo, 2006, 2008; Vrij, 2008). To overcome this problem, several credibility assessment methods have been developed to assist in lie detection. One group of methods is based upon the assumption that deception should be reflected in language. Criteria Based Content Analysis (CBCA; Steller & Köhnken, 1989), Reality Monitoring (RM; Johnson & Raye, 1981) and Scientific Content Analysis (SCAN; A Sapir, 1996) are well known examples.

Ample evidence exists showing that CBCA and RM perform above chance level at discriminating between truthful and fabricated accounts. Yet for SCAN only a few studies exist examining the diagnostic accuracy, and these studies typically fail to find an effect (see below, Nahari, Vrij, & Fisher, 2011; Porter & Yuille, 1996). In this study we examined to what extent CBCA, RM and especially SCAN can discriminate between truthful and fabricated accounts. In addition, we examined whether providing the participants with a detailed example statement increased the diagnostic accuracy of these methods.

CBCA, is based on the Undeutsch hypothesis, which states that statements based on true events differ in content and quality from statements that are the result of imagination (Undeutsch, 1967). The German psychologist Udo Undeutsch was the first to present his so-called reality criteria to aid in the credibility assessment of statements. Steller and Köhnken (1989) took these criteria a step further and generated a set of 19 different CBCA criteria. Even though CBCA was originally developed to assess the credibility of children's statements in cases of alleged sexual

abuse (Undeutsch, 1967), research has demonstrated that CBCA can also be used with adults (Blandon-Gitlin, Pezdek, Lindsay, & Hagen, 2009; Porter & Yuille, 1996; Steller & Köhnken, 1989). Research has shown that truthful statements are richer in these criteria than fabricated statements. Vrij (2005) investigated 37 studies about CBCA and concluded that CBCA is able to correctly classify 55% to 90% of the statements. The average classification accuracy in the investigated studies was 74.63%. In 2008, Vrij again investigated the accuracy for CBCA in 27 studies and found an average accuracy rate of 70.47%.

Reality Monitoring (RM) has also been employed to discriminate truthful from fabricated statements. Reality Monitoring originally refers to the mechanism that is used to distinguish between two types of memory, external and internal memories (Johnson & Raye, 1981). External memories are based on true experiences and are expected to contain sensory information such as details about smell, taste, sound and sight. In addition, true memories contain descriptions of context information and affective details. As a whole, they are expected to be sharp and vivid. In contrast, internal memories are based on imagined events and consequently consist of cognitive operations. These operations describe inferences (such as reasoning, search, decision, and imagery processes) made by the participants during the event or when describing the event at a later time. For example, “It appeared to me that she didn’t know the layout of the building” (Johnson, Raye, Foley, & Foley, 1981; Vrij, 2008). As a result, these imagined memories are less detailed compared to true memories. Masip, Sporer, Garrido, and Herrero (2005) reviewed the scientific RM deception literature and found that visual and auditory details, contextual information, time information and realism were the most discriminative criteria. Results about the cognitive operations were mixed. The RM classification was 68.80%, similar to that of CBCA (Vrij, 2008).

Scientific Content Analysis (SCAN) was developed by former Israeli polygraph examiner Avinoam Sapir (2005). Typically, a SCAN analysis starts with asking the suspect, witness, or alleged victim to write down ‘everything that happened’ during a particular time frame. This account is referred to as a ‘pure version’ of the event, as it has to be produced without the presence and interference of an investigator to minimize investigator influences. There is no standardized set of SCAN criteria, but a set of 12 criteria has been used in research (Bogaard, Meijer, Vrij, Broers, & Merckelbach, 2013; Nahari et al., 2011). Examples of criteria include whether pronouns are avoided or whether all individuals mentioned in the statement are introduced properly. The outcome of this analysis can be used to make a credibility judgement about the statement (A. Sapir, 2005). To date, SCAN was examined in only two experimental studies and neither experiment found evidence for SCAN. Porter and Yuille (1996) examined three SCAN criteria, structure of the statement; missing information and first person singular, past tense. Results indicated no differences between true and fabricated statements regarding these three criteria. In a more recent study, Nahari, Vrij and Fisher (2011) SCAN scores did not differ between truth tellers and liars. Interestingly, the same statements were also analyzed with RM and liars and truth tellers could be discriminated based on the RM results. Despite the lack of empirical evidence, SCAN is increasingly used for lie detection purposes in countries all over the world, such as Australia, Canada, Mexico, Israel, the Netherlands, Belgium, South Africa, UK, US, Qatar, and Singapore (Retrieved from <http://www.lscan.com/id29.htm>).

Research has shown that the accuracy of verbal credibility methods is sensitive to certain manipulations. For example, research investigating ‘coaching’ indicates that providing participants with specific information about the rationale and the different

criteria affects the CBCA and RM results (Caso, Vrij, Mann, & De Leo, 2006; Vrij, Akehurst, Soukara, & Bull, 2002, 2004; Vrij, Kneller, & Mann, 2000). Vrij, Kneller and Mann (2000) found that liars who were informed about CBCA criteria obtained higher CBCA scores than uninformed liars and that the CBCA scores of informed liars did not differ significantly from truth tellers' CBCA scores. Similar results were found regarding coached liars and RM scores (Vrij et al., 2004). The effect of coaching on SCAN evaluations has not been examined yet, but there is no theoretical reason why SCAN would not be vulnerable to coaching in the same way and to the same extent as CBCA and RM.

Besides coaching, researchers have also examined the influence of the type of interview on verbal credibility methods. For example, Köhnken, Schimossek, Aschermann and Höfer (1995) examined the effect of using the cognitive interview (CI) as an alternative to the structured interview on CBCA. Several studies have shown that the CI indeed led participants to give more correct details but also, to a lesser yet significant extent, more incorrect details (Fisher, Amador, & Geiselman, 1989; Geiselman, Fisher, MacKinnon, & Holland, 1986; Köhnken, Milne, Memon, & Bull, 1999; Memon, Meissner, & Fraser, 2010). Based on this memory enhancing effect, Köhnken et al. (1995) hypothesized that CI would influence the ability of CBCA to discriminate between true and fabricated accounts. Participants were asked to give a truthful or fabricated account of a blood-donation while being interviewed with either the structured interview or the CI. Results showed that statements of participants who were interviewed with the CI contained significantly more details misunderstood, lack of memory, self-doubts, unusual details and superfluous details than the statements of participants interviewed with the structured interview. Yet, CBCA scores did not differ between truthful and fabricated accounts depending on

the type of interview, showing that using the CI as an alternative to the structured interview did not influence the accuracy of the CBCA.

In the light of the methods' sensitivities to manipulations, in the present experiment we tested whether giving individuals an example of a detailed statement, fulfilling all relevant criteria of the CBCA, RM and SCAN, without explaining all the separate criteria would influence the methods' scores. Similar to Köhnken et al.'s (1995), the purpose of this example was to obtain a more comprehensive statement from participants. However, instead of asking memory-enhancing questions, we merely gave participants an example of the result we were interested in, i.e., a more detailed statement. Supplying individuals with a more detailed statement can yield different effects. It could result in liars obtaining higher scores on the different methods to the extent that their scores would not differ anymore from those of truth tellers, similar to the coaching effect. However, in the coaching experiments liars were informed how the veracity methods actually worked and which criteria to include, such specific instructions were not given in the present experiment. Without the exact knowledge of the different criteria within the example statement, it could become difficult for liars to produce a statement that is comparable with the example statement and of the same quality of a truth teller's statement. For truth tellers the example statement could encourage them to include more information. For them, the detailed example could act as a memory enhancing strategy, resulting in a more detailed statement. We therefore hypothesized that participants who received the example would give more comprehensive statements, resulting in higher scores on CBCA, RM and SCAN. In addition, we hypothesized that this effect would be more pronounced in truth tellers than in liars, thereby increasing discriminability of these methods. As mentioned earlier, besides investigating the effect of the example

statement, we also investigated to what extent these methods – and especially SCAN – are able to discriminate between true and false statements.

Method

Participants

A total of 64 participants (28 males), aged 17 to 27 years old ($M = 21.09$, $SD = 2.27$) took part in the study. Mother tongue Dutch was a prerequisite. This study was approved by the Ethics Committee of the faculty of Psychology and Neuroscience of Maastricht University.

Materials

Verbal credibility methods

Criteria included in the example statement

Although CBCA originally consists of 19 criteria, only nine criteria were included in the example statement. We left out the motivational-related criteria, and the specific criteria “unexpected complications during the incident” and “related external associations” because of weak empirical support in previous studies (Vrij, 2005).

Furthermore, the criterion ‘accurately reported details misunderstood’ was not suitable for the current study as it refers to typical child witness issues. Finally, three criteria were omitted because they had contradicting interpretations depending on the method used. These particular criteria were ‘Spontaneous corrections’ (both CBCA and SCAN), ‘Admitting lack of memory’ (CBCA) vs. ‘Lack of conviction or memory’ (SCAN) and ‘Unstructured production’ (CBCA) vs. ‘Out of sequence and extraneous information’ (SCAN). For CBCA the presence of these criteria indicates truthfulness while for SCAN the presence of these criteria indicates deception. As a

result, the following nine CBCA criteria were of use to this study (for a complete description of all criteria see Appendix A): (1) Logical structure, (2) Quantity of details, (3) Contextual embedding, (4) Descriptions of Interactions, (5) Reproduction of conversation, (6) Unusual details, (7) Superfluous details, (8) Accounts of subjective mental state and (9) Attribution of perpetrator's mental state (Steller & Köhnken, 1989; Vrij, 2008). All criteria are expected to be more present in truthful statements than in deceptive statements.

We wanted to create an example statement that would be perceived as highly credible. We therefore only included in the statement criteria that indicate truthfulness and left out criteria that indicate deception. For the RM method this means we excluded the criterion "cognitive operations". As a result, seven criteria were included in the example statement: (1) Clarity, (2) Perceptual information, (3) Spatial information, (4) Temporal information, (5) Affect, (6) Reconstructability of the story, and (7) Realism. All seven criteria are expected to be more present in truthful statements. (Sporer, 1997; Vrij, 2008).

For the SCAN method, 6 instead of the 12 criteria reported by Vrij (2008) were used. Three overlapping criteria mentioned above were excluded, as well as the criterion "Denial of allegations" as – given the nature of the statements - it is not appropriate for the current study (participants were not accused of anything and therefore had no reason to deny the allegations). Furthermore, as mentioned before, we also excluded the criteria indicating deception, as they could not be incorporated in the example statement. The resulting six criteria included were (1) Social introduction, (2) Structure of the statement, (3) Emotions, (4) Objective and subjective time, (5) First person singular, past tense, and (6) Pronouns. See appendix A for the complete lists and definitions of the criteria.

Example statement

The example statement was produced by one additional participant who read all the necessary literature about CBCA, RM and SCAN and received a 1.5 hour training in the three methods. To make sure the statement was detailed enough, the participant was first interviewed about the event, a recent surprise helicopter flight in New York. Next, the participant was asked to write down a very detailed (true) statement about the event, and to avoid including criteria that indicate deception. The statement consisted of 1291 words, which is approximately three pages in writing, and aimed to fulfill all criteria of CBCA, RM and SCAN indicating truth, described above.

To check whether each separate criterion indicating truthfulness was indeed present in the example statement, a pilot study with 12 participants (7 males, $M = 20.75$ years, $SD = 2.45$ years) was conducted. Four participants analyzed the example statement with CBCA, four with RM and four with SCAN. All participants received a 1 hour training about the method they were going to use, and scored all criteria indicating truthfulness on a 3-point scale including 0 (not present), 1 (present), and 2 (strongly present; Blandon-Gitlin et al., 2009; Gödert, Gamer, Rill, & Vossel, 2005). As we instructed the writer of the story not to include criteria indicating deception, we did not further investigate these criteria.

Table 1 shows an overview of the different scores for each criterion separated for each method. For each method we calculated the total method score averaged for the four raters. For CBCA, the four raters obtained an average total score of 11.75. As can be seen in Table 1, criterion (9) “Attribution of perpetrator’s mental state” was scored as absent in the example statement. This is due to the positive nature of the statement, where no perpetrator was present. For RM, the four raters obtained an average score of 10. None of criteria were scored as absent. Finally, for SCAN, the

four raters obtained an average score of 8. The criterion (4) “objective and subjective time” was scored as present (1) by two raters, and as absent by the other two. We believe that these scores indicate that the different criteria were sufficiently presented in the example statement.

Procedure

Participants were told that the study was about the discriminability of verbal lie detection methods. Participants were asked to think for maximum three minutes about a real and a fabricated story. They were told that both stories should be about a recent negative event in their life, either emotionally or physically. It was stressed that the fabricated story should consist of a realistic scenario that never actually occurred in their life. Furthermore, half of the participants received the ‘example statement’, and were told that they had to produce a statement in similar style to the example statement about the event they actually experienced and about an event they completely fabricated. The other half of the participants were merely asked to produce the two statements as detailed as possible. All participants were told that their statements would be used to test deception detection methods and that they had to try to write their stories as convincing as possible. The order in which the participants wrote the two statements was counterbalanced. Participants were reassured that their stories would be treated confidentially and anonymously. No time limit was given for the production of the statements.

Statement coding

After participants finished writing their statements, the statements were analysed with SCAN, RM and CBCA by two independent raters for each method. All six raters were

females with an average age of 22.8 years. The raters were made familiar with the literature discussing one of the three different verbal analyzing methods they were expected to use (Masip et al., 2005; Nahari et al., 2011; Steller & Köhnken, 1989). Participants also received the appropriate pages of the different chapters of Vrij (2008) about CBCA (chapter 8; 207-213), or RM (chapter 9; 266-269) or SCAN (chapter 10; 282-287). In addition, the raters received a 1.5-hour training about the method they were expected to use. During the training all the criteria were discussed separately and examples were given. After all criteria were discussed, raters received two practice statements and these were also discussed. After this training was completed, the coders started analyzing the 128 written statements of this study. Two raters analyzed the statements with CBCA, two raters with RM and two raters with SCAN. Raters coded the presence of all criteria, including those that were excluded from the example statement ('Spontaneous corrections', 'Admitting lack of memory', and 'Unstructured production' for CBCA, 'Cognitive operations' for RM, and 'Lack of conviction or memory', 'Out of sequence and extraneous information', 'Out of sequence and extraneous information', 'Missing information', and 'Change of language' for SCAN). Thus, for CBCA the raters used used a list of 12 criteria. For RM 8, and for SCAN 11.

Again, all criteria were scored on a 3-point scale ranging from 0 (not present) to 2 (strongly present). However, criteria indicating deception, indicated with an asterisk in the appendix, were scored in reverse, ranging from -2 (strongly present) to 0 (not present).

Inter-rater reliability

Inter-rater reliability for all rated criteria was calculated by means of Cohen's Kappa for each separate criterion per method. For CBCA, values varied between 0.47 to 0.91 with an average Cohen's Kappa of 0.70. For RM, values varied between 0.32 and 0.90 with an average Cohen's Kappa of 0.53. Finally, for SCAN, values varied between 0.49 and 0.94 with an average Cohen's Kappa of 0.72. These results indicated that there is moderate (RM) to high agreement (CBCA, SCAN) between the two raters. Although the scores are sometimes low for individual criteria, we do not think this is too damaging as we only used total scores in our analyses. Previous studies in lie detection have used similar reliabilities, and also lower reliabilities than the ones found in the current study (Blandon-Gitlin et al., 2009; Vrij, 2005).

Results

Statement Type

Statements were coded as either emotional or physical. A statement was coded as physical when the writer wrote about a physical injury that s/he experienced her/himself. All other statements were coded as emotional. For example, a statement describing the grandpa being treated for cancer would be coded as emotional. A statement describing a car crash in which the writer was hurt would be coded as a physical statement. For the true statements, 53 statements discussed an emotional event, and 11 statements discussed a physical event. For the fabricated statements, 50 statements consisted of an emotional event, and 14 statements discussed a physical event. This difference was not significant [$\chi^2(1, N=64) = \text{n.c.}, p=0.66$], indicating no preference for a certain type of statement when lying or telling the truth. Furthermore, we also correlated the physical/emotional score for the true statement with that of the fabricated statement. Results showed a non-significant correlation ($r = -0.04$)

suggesting no clear evidence of either shifting between or sticking to a specific statement type.

Verbal Credibility Methods

Diagnostic accuracy

To exclude any potential effect of the example statement, the analyses regarding the diagnostic accuracy of the three methods were only carried out on the control group. For CBCA, the raters coded 12 criteria. The other seven criteria of CBCA were excluded because of weak empirical support or because they were not suitable for the current study. For RM the raters coded all 8 criteria. For SCAN, due to the nature of the statements, the criterion ‘Denial of allegation’ was not coded, resulting in 11 included criteria.

We calculated the total score for each method, by summing up 12 criteria for CBCA, 8 criteria for RM and 11 for SCAN. True statements included significantly more CBCA criteria ($M = 7.19, SD = 2.09$) than fabricated statements ($M = 6.17, SD = 2.16$) [$t(31) = 2.21, p = 0.04$]. True statements also included significantly more RM criteria ($M = 5.70, SD = 2.03$) than fabricated statements ($M = 4.53, SD = 2.10$) [$t(31) = 2.53, p = 0.02$]. However, SCAN scores did not differentiate between true ($M = 4.81, SD = 1.63$) and fabricated statements ($M = 4.62, SD = 1.94$) [$t(31) = 0.50, p = 0.62$]. To allow comparison with other studies, we calculated Cohen’s d for the main effect of veracity for the different methods. The d -values were 0.48 for CBCA, 0.56 for RM, and 0.11 for SCAN.

A non-significant effect for the SCAN total score does not rule out that some individual SCAN criteria discriminated significantly between truth tellers and liars.

To further examine the SCAN criteria. Table 2 shows that not a single SCAN criterion significantly discriminated between true and fabricated statements.

- Please insert Table 2 about here -

Effect of the example statement

To investigate to what extent the example statement influenced the number of criteria present in the statements, we only included the criteria that were included in the example statement in this analysis. Total CBCA score was calculated by summing up the 9 CBCA criteria, total RM score by summing up the 7 RM criteria, and total SCAN score by summing up the 6 SCAN criteria that were present in the example statement (see methods section). Mean scores and standard deviations of the three methods are presented in Table 3. A 2 (Veracity: truth vs. fabricated) x 2 (Example: yes vs. no) mixed ANOVA on the total CBCA scores revealed a significant main effect for Veracity [$F(1, 62) = 12.52, p < .001, \eta_p^2 = 0.17$] with true statements receiving a higher CBCA score than fabricated statements. Results also showed a significant main effect for Example [$F(1, 62) = 19.65, p < .001, \eta_p^2 = 0.24$] with participants who received an example statement writing a statement that scored higher on CBCA. Results did not show a significant Veracity X Example interaction [$F(1, 62) = .23, p = .63, \eta_p^2 = 0.01$] indicating that the example statement did not significantly influence the discriminability of the CBCA.

For RM, the results of the mixed ANOVA showed a significant main effect for Veracity [$F(1, 62) = 11.71, p < .001, \eta_p^2 = 0.16$], with true statements scoring higher on RM than fabricated statements. Results also showed a significant main effect for Example [$F(1, 62) = 24.52, p < .001, \eta_p^2 = 0.28$] with participants who received an example writing a statement that scored higher on RM. Again, results did not show a

significant interaction between Veracity and Example [$F(1, 62) = .15, p = .70, \eta_p^2 = 0.002$] indicating that the example statement did not significantly influence the discriminability of the RM.

For SCAN, the mixed ANOVA showed no significant main effect for Veracity [$F(1, 62) = 0.35, p = .556, \eta_p^2 = 0.006$], revealing that SCAN did not accurately discriminate between true and fabricated statements. However, results did show a significant main effect for Example, [$F(1, 62) = 7.15, p = .001, \eta_p^2 = 0.10$], indicating that the example statement caused participants to receive a higher SCAN score. Finally, results showed no significant interaction effect between Veracity and Example [$F(1, 62) = .005, p = 0.94, \eta_p^2 = 0$] indicating that the example statement did not significantly influence the discriminability of the SCAN scores.

-Please insert Table 3 about here-

Number of words

Means and standard deviations of the number of words present in the statements are presented in Table 4. Participants in the example condition wrote on average 296.78 ($SD = 98.76$) words, while participants in the control condition wrote on average 181.94 ($SD = 55.75$) words. A 2 (Veracity) x 2 (Example) mixed design analysis showed no main effect for Veracity, [$F(1, 62) = 2.01, p = .16, \eta_p^2 = 0.03$]. However, results did show a significant main effect for Example, $F(1, 62) = 37.38, p < .001, \eta_p^2 = 0.38$, indicating that participants in the example statement condition wrote a longer statement compared to the control group. Results showed no significant interaction between Veracity and Example, $F(1, 62) = .003, p = .96, \eta_p^2 = < 0.01$, meaning that

receiving an example statement influenced true and false in a similar way (see Table 3).

-Please insert Table 4 about here-

Discussion

The aim of the present study was to test the diagnostic accuracy of CBCA, RM and SCAN. Furthermore, we investigated the effect of supplying participants with a detailed example statement on verbal credibility assessment scores. The results showed that both CBCA and RM scores were significantly higher for true compared to false statements. In contrast, SCAN failed to discriminate between true and false statements. Further investigation of the separate SCAN criteria also failed to find supporting evidence for its diagnostic accuracy.

These findings are in line with the previous studies of Nahari et al. (2011) and Porter & Yuille (1996) that failed to find support for SCAN. Nahari et al. (2011) also investigated both a SCAN total score and compared these scores to RM total scores. As in our study, RM was able to accurately discriminate between true and fabricated statements, while SCAN performed poorly in discriminating between these statements. Although the lack of significant findings for SCAN is in line with previous research, one limitation deserves attention. First, as part of their training, the SCAN raters received the article of Nahari et al. (2011), which concluded that SCAN was not able to accurately discriminate between true and false statements. This information may have influenced the motivation of the raters and thereby the quality of the analysis of the statements.

Supplying participants with an example statement resulted in higher CBCA, RM and SCAN scores for both fabricated and truthful accounts, failing to increase

diagnostic accuracy. This lack of increased diagnostic accuracy as a result from the example statement for CBCA are in line with Köhnken et al. (1995) who showed that using CI resulted in an increased use of certain CBCA criteria. Supplying an example statement affected both true and false statements in a similar way. Meaning, the use of an example statement did not influence the discriminability of CBCA, RM or SCAN. Participants who received the example statement produced longer statements than participants who did not receive the example. As such, although an example statement did not improve the ability to detect deceit, providing suspects in real life with an example may still be useful as a statements produced after receiving an example are likely to contain more information. As obtaining information from suspects is often seen as the core of interviewing (Bull, 2010; Fisher, 2010), providing an example statement may contribute to this. The obtained information can in turn be checked and validated and may provide more insight into the suspect's involvement in the crime.

To maintain ecological validity, we deliberately did not explain the criteria to the participants (as is done in coaching studies), but hypothesized that truth tellers would implicitly incorporate more criteria from the example statement in their statement than liars. Importantly, besides the main effect on the number of words, there was also a main effect for all three method scores, showing that participants who received the example statement did include more criteria in their statement. Thus, even though the participants may not have been explicitly aware of the criteria, this shows that our example statement manipulation was successful and that the participants did not simply write a longer statement.

In the current study, participants were free to choose the events on which they based their statements. This resulted in a variety of events described, and as a consequence the veracity of the statements (i.e., ground truth) could not be verified. It

cannot be ruled out that some of the participants failed to follow instructions, which in turn could have influenced our results. Although it could be argued that this is a limitation of the design of this study, it is also important to realize that such a variety of events described is in fact what investigators are faced with in practice. Thus, it can also be argued that by trading some degree of control such a design increases external validity.

The findings of the current study have two practical implications. First, as our findings replicate that SCAN criteria do not discriminate between true and fabricated statements, the extensive use of SCAN in practice should be discouraged. Second, our data shows that providing participants with an example statement results in higher CBCA, RM, and SCAN scores for both truthful and fabricated statements. This means that when using such a manipulation in real life, one should be cautious about veracity verdicts based on a high score, as these high scores may simply be due to the example statement.

In conclusion, results clearly indicated that the lie detection methods differ in their sensitivity to measuring truthfulness as RM and CBCA, but not SCAN were able to accurately discriminate between true and fabricated statements. Supplying individuals with a detailed example statement did not increase or decrease discriminability of the verbal credibility assessment, but did increase the length of the statement, which might be useful for investigative purposes.

References

- Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner, 15*, 6-11.
- Blandon-Gitlin, I., Pezdek, K., Lindsay, S., & Hagen, L. (2009). Criteria-based Content Analysis of true and suggested accounts of events. *Applied Cognitive Psychology, 23*, 901-917.
- Bogaard, G., Meijer, E., Vrij, A., Broers, N. J., & Merckelbach, H. (2013). SCAN is largely driven by 12 criteria: Results from field data. *Psychology, Crime and Law*. doi: 10.1080/1068316X.2013.793338
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Individual Differences, 10*, 214-234.
- Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*, 477-492.
- Caso, L., Vrij, A., Mann, S., & De Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology, 11*, 99-111.
- Fisher, R. P., Amador, M., & Geiselman, R. E. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology, 74*, 722-727.
- Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *American Journal of Psychology, 99*, 385-401.
- Gödert, G. W., Gamer, M., Rill, H., & Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal and Criminological Psychology, 10*, 225-245.

- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *The American Journal of Psychology*, 94, 37-64.
- Köhnken, K., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, 5, 3-27.
- Köhnken, K., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, 80, 671-684.
- Masip, J., Sporer, A. L., Garido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime and Law*, 11, 99-122.
- Memon, A., Meissner, C. A., & Fraser, J. (2010). The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, 16, 340-372.
- Nahari, G., Vrij, A., & Fisher, R. P. (2011). Does the truth come out in the writing? SCAN as a lie detection tool. *Law and Human Behavior*, 1-11.
- Porter, S., & Yuille, Y. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context *Law and Human Behavior*, 20, 443-458.
- Sapir, A. (1996). *The LSI course on scientific content analysis (SCAN)*. Phoenix, AZ: Laboratory for Scientific Interrogation.
- Sapir, A. (2005). *The LSI course on scientific content analysis (SCAN)*. Phoenix, AZ: Laboratory for Scientific Interrogation.

- 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., & Köhnken, G. (1989). Criteria Based Statement Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von Aussagen. In U. Undeutsch (Ed.), *Handbuch der psychologie Vol 11: Forensische Psychologie*. Göttingen, Germany: Hogrefe.
- Vrij, A. (2005). Criteria Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*, 3-41.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: Wiley.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior, 26*, 261-283.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and Reality Monitoring scores as a function of age, coaching and deception. *Canadian Journal of Behavioural Science, 36*, 113-126.
- 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.

Appendix A (adapted from Vrij (2008))

Criteria Based Content Analysis (CBCA)

- (1) Logical structure: The statement should be coherent and does not contain logical inconsistencies or contradictions
- (2) Quantity of details: The statement should be rich in detail and include specific descriptions of place, time persons, objects and events.
- (3) Contextual embedding: The events should be placed in time and locations, and the actions are connected with other daily activities and/or customs.
- (4) Descriptions of Interactions: The statement should contain information that interlinks at least the alleged perpetrator and witness.
- (5) Reproduction of conversation: The statement should report parts of the conversation in the original form or the different speakers are recognizable in the reproduced dialogues.
- (6) Unusual details: Refers to details of people, objects, or events that are unique, unexpected or surprising but meaningful in the context.
- (7) Superfluous details: Refers to details that are in connection with the allegations that are not essential for the accusation.
- (8) Accounts of subjective mental state: Refers to the witness describing the development and changes of his or her feelings at the time of the incident.
- (9) Attribution of perpetrator's mental state: Refers to the witness describing the perpetrator's feelings, thoughts, or motives during the incident.
- (10) Unstructured production: The information in the statement is presented in a non-chronological order.

- (11) Spontaneous corrections: The statement contains corrections or information is added to previously provided information, without any interference of the interviewer.
- (12) Admitting lack of memory: the statement contains information admitting lack of memory such as “I don’t know” or “I don’t remember what happened after he hit me”.

Reality Monitoring (RM)

- (1) Clarity: Refers to the clarity and vividness of the statement.
- (2) Perceptual information: Refers to the presence of sensory information in a statement (sound, smells, tastes, physical sensations and visual details.
- (3) Spatial information: Refers to information about locations or the spatial arrangement of people and/or objects.
- (4) Temporal information: Refers to information about when the event happened or explicitly describes a sequence of events.
- (5) Affect: Refers to information that describes how the participant felt during the event.
- (6) Reconstructability of the story: Examines whether it is possible to reconstruct the event on the basis of the information given.
- (7) Realism: Examines whether the story is plausible, realistic and makes sense.
- (8) Cognitive operations: Refers to descriptions of inferences made by the participant ant the time of the event.*

Scientific Content Analysis (SCAN)

- (1) Social introduction: Refers to how the persons described in the statement are introduced.
- (2) Structure of the statement: Refers to the balance of the statement. In a truthful statement 20% is used to describe activities leading up to the event, the next 50% to describe the actual event, and the final 30% to discuss what happened after the event.
- (3) Emotions: Refers to whether there are emotions described in the statement.
- (4) Objective and subjective time: Refers to how different time periods are covered in the statement. Objective time refers to the actual duration of events described, whereas subjective time refers to the amount of words spent to describe these events.
- (5) First person singular, past tense: Refers to the format in which a statement is written. Deviations from first person singular, past tense can indicate deception.
- (6) Pronouns: Refers to the use of pronouns in the statement. Omitting pronouns suggests reluctance of the writer to commit him/herself to the described actions.
- (7) Spontaneous corrections: Refers to all corrections that are made within the statements.*
- (8) Lack of conviction or memory: Refers to when the writer is vague about certain elements within the statement (e.g. "I think...", "I guess...") or when the writer admits he or she has forgotten something. *

- (9) Out of sequence and extraneous information: Examines whether the statement recounts the events in chronological order and whether there is extraneous information that does not seem relevant.*
- (10) Missing information: Refers to phrases in the statement that indicate that some information has been left out (e.g., finally, later on).*
- (11) Change in language: Refers to the change of terminology or vocabulary in the statement. A change in language indicates that something has altered in the mind of the writer. *

Table 1.

Detailed overview of the scores from the different raters for the pilot statement

Method	Participant	Criteria								
		c1	c2	c3	c4	c5	c6	c7	c8	c9
CBCA	1	2	2	2	1	2	2	0	1	0
	2	2	2	1	1	2	1	2	2	0
	3	2	2	2	1	1	0	1	2	0
	4	1	2	2	1	2	0	1	2	0
RM	5	2	2	2	2	1	1	2	-	-
	6	2	1	1	1	2	1	2	-	-
	7	2	1	2	2	2	1	2	-	-
	8	2	2	2	2	2	1	2	-	-
SCAN	9	1	2	2	0	2	2	-	-	-
	10	1	2	2	1	2	2	-	-	-
	11	1	2	2	0	2	2	-	-	-
	12	0	2	2	1	2	2	-	-	-

Note. The numbers of the criteria refer to the numbers of the criteria mentioned in the Appendix A. For each method, the numbers refer to different criteria.

Table 2.

Detailed overview of the Means and Standard Deviations for each SCAN criterion as a function of veracity

Criteria	True		Fabricated		Cohen's d
	M	SD	M	SD	
Social Introduction	0.70	0.71	0.83	0.63	-0.19
Spontaneous corrections	-0.42	0.58	-0.38	0.55	-0.07
Lack of conviction or memory	-0.03	0.18	0	0	-
Structure of the statement	0.78	0.61	0.63	0.62	0.24
Emotions	0.70	0.58	0.69	0.64	0.02
Objective and Subjective time	0.30	0.42	0.31	0.42	-0.02
Out of sequence and extraneous info	-0.25	0.38	-0.25	0.38	0
Missing information	-0.25	0.40	-0.27	0.44	0.05
First person sing, past tense	1.70	0.54	1.55	0.53	0.28
Pronouns	1.67	0.45	1.72	0.46	-0.09
Change in language	-0.09	0.24	-0.2	0.42	0.33

Table 3.

Means and Standard Deviations for total scores on CBCA, RM and SCAN as a function of condition for the criteria included in the example statement

Condition	CBCA		RM		SCAN	
	True	False	True	False	True	False
Explanation	8.50 (2.71)	7.27 (1.81)	7.77 (2.42)	6.88 (2.04)	6.67 (1.55)	6.56 (1.58)
Control	6.48 (1.77)	5.55 (1.90)	5.72 (2.04)	4.56 (2.08)	5.85 (1.23)	5.72 (1.61)
Total	7.49 (2.49)	6.41 (2.04)	6.76 (2.45)	5.72 (2.35)	6.27 (1.44)	6.14 (1.63)

Note. CBCA, RM and SCAN are abbreviations referring to the methods explained in Appendix A.

Table 4.

Means and Standard deviations for the number of words in true and fabricated statements for both conditions

Condition	True		Fabricated	
	M	SD	M	SD
Explanation	301.78	102.66	291.78	96.08
Control	187.34	58.79	176.53	52.91
Total	244.56	101.06	234.16	96.40