

SCAN is largely driven by 12 criteria: Results from sexual abuse statements.

Glynis Bogaard¹, Ewout H. Meijer¹, Aldert Vrij², Nick J. Broers³ and Harald Merckelbach¹

¹ *Clinical Psychological Science, Maastricht University, the Netherlands*

² *Psychology, University of Portsmouth, the United Kingdom*

³ *Methodology and Statistics, Maastricht University, the Netherlands*

Glynis Bogaard*

Ewout H. Meijer

Harald Merckelbach

Department of Clinical Psychological Science

Maastricht University

P.O. Box 616, 6200 MD Maastricht

The Netherlands

Aldert Vrij

Department of Psychology

University of Portsmouth

Portsmouth

The United Kingdom

Nick J. Broers

Department of Methodology and Statistics

Maastricht University

P.O. Box 616, 6200 MD Maastricht

The Netherlands

*Corresponding author. Email: Glynis.bogaard@maastrichtuniversity.nl

2 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Abstract

Scientific Content Analysis (SCAN) is increasingly used by investigative authorities to evaluate the credibility of statements made by witnesses and suspects. SCAN, however, lacks a well-defined list of criteria, and does not involve a standardized scoring system. In the current study, we investigated which SCAN criteria are represented in actual statements. To this end, we analysed 82 sexual abuse cases of the Dutch police in which SCAN had been applied. Two independent coders scored the presence of various SCAN criteria in the (i) written statements from victims, suspects, and witnesses, and also looked at the (ii) recommendations for follow-up investigations that were derived from SCAN. Results showed that SCAN is primarily driven by 12 criteria. Results also indicated a low inter-rater agreement for most SCAN criteria, suggesting SCAN is insufficiently developed as a forensic tool. Still, the 12 criteria can be used as a starting point for future research on their psychometric properties.

Keywords: lie detection; Scientific Content Analysis; verbal cues, SCAN.

3 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

SCAN is largely driven by 12 criteria: Results from sexual abuse statements.

Investigative authorities are often confronted with deceptive suspects, with witnesses who raise dubious claims, and with bogus victims who fabricate traumatic stories (Greer, 2000; Gudjonsson, Sigurdsson, Asgeirsdottir, & Sigfusdottir, 2007; Lisak, Gardinier, Nicksa, & Cote, 2010). Meanwhile, research has shown the detection of deception to be challenging. Experts, including police officers, generally perform only just above chance level when they base their judgements about deceit on the verbal and nonverbal behaviour of suspects or witnesses (Aamodt & Custer, 2006; Bond & DePaulo, 2006; Vrij, 2008b). Although there is good evidence that verbal indicators of deceit are generally more diagnostic than non-verbal indicators, people intuitively believe that invalid – non-verbal - cues indicate deception (Akehurst, Kohken, Vrij, & Bull, 1996; Strömwall & Granhag, 2003; Vrij, 2008a; Vrij, Granhag, & Porter, 2010).

With this in mind, some researchers (e.g., Vrij, 2008b) have argued that to optimize credibility assessment of suspects, witnesses, and victims, investigative authorities should rely on verbal rather than non-verbal cues to deception. One of the tools that is being used for this purpose is Scientific Content Analysis (SCAN; Sapir, 2005). SCAN was developed in the mid eighties by former Israeli polygraph examiner Avinoam Sapir. Over the past few years, SCAN has been used by police investigators in Belgium, Canada, Israel, Mexico, the Netherlands, Singapore, South-Africa, the United Kingdom, and the United States (Vrij, 2008a). Furthermore, SCAN is used by Federal Agencies (including CIA), Military Law Enforcement (including US Army, US Air Force, US Marine), Private Corporations, and Social Services (retrieved from www.lsiscan.com/id29.htm). Typically, a SCAN analysis starts with asking the suspect, witness, or alleged victim to write down ‘everything that happened’ during a

4 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

critical period of time. Sapir (2005) refers to this as the ‘pure version’ of the event, produced without any interference from a police officer. Next, this pure version is matched against criteria such as the extent to which there are gaps in the chronology or the extent to which pronouns are avoided. The outcome of this analysis can be used to make a veracity judgement about the statement and/or to guide subsequent interrogations (see for a more detailed description: Vrij, 2008a).

SCAN has been criticized for lacking a well-defined list of criteria, as well as a standardized scoring system. Police officials using SCAN typically attend a three-day course, but, as the first author discovered while attending this course, no checklist of criteria is offered, nor does the SCAN manual contain such a list. The absence of well-defined SCAN criteria is reflected in the literature, as different authors report different sets of criteria when describing SCAN. For example, in his field study, Driscoll (1994) mentions 10 SCAN criteria, the Belgian police superintendent Bockstaele describes either 14 (Bockstaele, 2008a) or 16 criteria (Bockstaele, 2008b), while in her research governed by the UK’s Home Office, Smith (2001) reported 12 criteria. SCAN has also been used in the Netherlands during a pilot by the vice squad of police AMS. In this pilot, the police employed 21 criteria derived from the SCAN course manuals. Listing all linguistic features of SCAN that are mentioned in the literature results in as many as 28 distinct criteria (see Table 1 and Appendix A). Moreover, different studies rely on different interpretations of the criteria. For example, Smith’s definition of “unnecessary connections/missing information” overlaps with two distinct SCAN criteria - namely “connections” and “missing and/or unnecessary links” - mentioned by Driscoll (1994). It is important to note that despite focusing on a range of criteria, none of the above mentioned studies found any supporting evidence for SCAN as a lie detection tool.

5 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

- Please insert Table 1 about here -

Given the lack of standardization, it is not surprising that the reliability of SCAN is low. Smith (2001) analysed levels of agreement between eight SCAN analysts, divided into three groups based on their SCAN experience (experienced users, occasional users, and infrequent users) in how they applied SCAN criteria to suspect statements. All analysts were given 27 statements from actual cases and asked to analyse these statements with SCAN. The inter-rater reliability was found to be disappointingly low. Specifically, analysts frequently relied on different criteria when coding the same statement. The highest degree of inter-rater consistency was found for the 'pronouns' criterion, yet this criterion still only achieved a 40% level of agreement. That is, the 'pronouns' criterion was scored to be present in 10 out of 27 statements by one or more groups, while in only 4 out of these 10 statements all groups agreed on the presence of the 'pronouns' criterion. Smith's (2001) study showed that analysts judged many statements as deceptive, but their judgments were based on the use of different criteria.

The present study had two aims. First, to examine which of the various SCAN criteria are present in actual sexual abuse cases statements. We analysed the statements made by suspects and alleged victims in 82 sexual abuse cases that were provided to us by the Dutch police. These statements were part of a pilot conducted by the vice squad of the Amsterdam (AMS) Police in which SCAN was applied to statements written by suspects or alleged victims in criminal cases. We examined which SCAN criteria were identified by the Amsterdam police SCAN analysts in these written statements. To obtain further insight into which criteria are used by

6 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Amsterdam police, we examined which of the SCAN criteria were included by the SCAN analysts in their recommendations for follow-up forensic investigations (for details, see method). Although the current study only relied on sexual abuse statements, the results of this study may still benefit research on SCAN because it would allow for a more standardized list of criteria that can be used in future research. Second, we calculated the inter-rater reliability of the AMS police SCAN analysts to obtain insight into the reliability of the SCAN method in real life practice, as sufficient inter-rater agreement is an important issue for any assessment method. Thus, the current study addresses reliability rather than validity issues related to the SCAN.

Method

The Amsterdam Police Pilot

In the vice squad of the Amsterdam police conducted a pilot, using SCAN in 115 sexual abuse cases. Alleged victims, witnesses, and suspects were asked to write down their version of what had happened, upon arrival at the police station, before any interrogation or interview. These statements were then submitted to a SCAN analysis (for details see below), and the results of this analysis were used to guide subsequent interrogation. The goal of this pilot was to explore to what extent SCAN might help in determining the credibility of victims, witnesses, and suspects.

The pilot team consisted of four SCAN analysts (three women), who all had completed the basic SCAN course (see www.lsiscan.com) a few months before the pilot started. At the start of the project, none of the SCAN analysts were experienced in employing SCAN in practice. During the pilot project, they frequently consulted the developer of the instrument (i.e., Avinoam Sapir). The police officers who

7 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

analysed the statements will be referred to as SCAN analysts throughout the entire manuscript.

All statements in the pilot were analysed by two of the four trained SCAN analysts. In a typical SCAN analysis, the criteria in a statement are first colour coded. Table 1 gives a detailed overview of how the SCAN analysts coded each criterion within the written statements. For example, every person who is mentioned in the statement is highlighted in green (Appendix A, criterion 1). For some criteria, their presence is simply indicated in the margin. For example, the structure of the statement is written in the margin at the end of the statement (Appendix A, criterion 5). Finally, a subset of the criteria is not coded in the statement at all, and is only described in the report about the statement (e.g., features like first person singular, past tense; Appendix A, criterion 2). Having analysed the material in this way, each analyst wrote a report about the statement. Based on these reports, a final report was produced that combined all the individual analyses and recommendations. Such a final report would include information about specific questions that could be asked during a subsequent interview based on the SCAN analysis. For example, when the SCAN analysis reveals several unasked explanations within the statement during a specific time frame, a possible recommendation within the report could be: “Ask further explicit questions about the specific time frame”. As another example, the writer may not properly introduce an important person within the statement. According to SCAN this could indicate a difficult relationship between the writer and the introduced person, and a recommended question could be: “Tell us about the relationship with [incomplete introduced person]”.

Thus, for each case, the SCAN record, as coded by two SCAN analysts, consisted of:

8 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

- i. an unprocessed copy of the original written statement of the interviewee;
- ii. copies of the original statement including the coding scheme of the SCAN analysts (2 versions);
- iii. individual reports about the statement written by each analyst (2 reports);
- iv. a final report that combined the individual analyses and recommendations.

Cases included in the current study

Of the 115 cases that were part of the pilot, 11 could not be used because they were incomplete: nine did not contain a SCAN coded statement, one did not contain a SCAN report with recommendations, and one record consisted of unreadable handwriting. Eighty-two statements were collected by asking alleged victims ($n = 43$), suspects ($n = 18$), and witnesses ($n = 21$). The remaining cases did not contain a written statement, but consisted of other material such as letters from victims ($n = 6$), suspects ($n = 7$) and witnesses ($n = 7$), and e-mails from two witnesses. In accordance with the SCAN guidelines that emphasize the pure narrative version as a starting point for analysis, and to keep the set of statements homogeneous, these cases were also omitted. This resulted in a set of 82 cases that were included in the analysis. All cases were sexual abuse cases.

Analysis

Present/absent

SCAN criteria in the written statements and in the final report containing the recommendations were coded separately. To identify the presence of the SCAN criteria in the written statements, two independent coders evaluated whether each of the 21 criteria that were used by police AMS were present within the statements (see

9 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Table 1 and Appendix A for definitions). One coder followed the SCAN basic course and the other coder read the SCAN course manual and was familiar with the SCAN literature and coding scheme. Thus, these coders did not analyse the statements themselves, but merely investigated the coding scheme that the SCAN analysts had produced for each statement. The evaluation by these coders will be referred to as SCAN evaluation. Presence of a criterion in the coding scheme for a statement was coded as '1' and absence as '0'. Criteria that were coded as present had to be either highlighted within the statement or noted in its margin, resulting in a 0/1 coding per criterion, per statement.

To investigate which of the SCAN criteria contributed to recommendations for follow-up investigations or interviews, the same two independent coders coded whether each of the 21 criteria used by police AMS were present ('1') or absent ('0') in either the individual reports or the final report. Again, this resulted in a 0/1 coding per criterion, per statement.

Frequency

The 0/1 coding described above, however, is suboptimal for situations where a criterion is scored by both analysts, but at different locations within the statement. To obtain more detailed data in the presence of SCAN criteria, we also investigated the exact frequency of the different criteria within the written statements. This frequency analysis was limited to the written statement, because the recommendations are reports about the criteria, but do not contain the criteria themselves. Consequently, the frequency count will have no additional value over the presence (1) or absence (0) coding of the recommendations.

To identify the exact frequency of the SCAN criteria in the written statements, the same independent coders who coded the presence or absence of the criteria also

10 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

coded how many times each of the 21 criteria employed by police AMS were present within each statement according to each SCAN analyst. Criteria that were coded as present had to be either colour coded within the statement or noted in its margin. Each coder gave a frequency count of each criterion for each line of the written statements. For example, when a SCAN analyst highlighted two social introductions in line 1 and one introduction in line 6, the total frequency of this criterion would be 3. Furthermore, the coding schemes were compared to make sure that when both coders coded the same number of criteria within one line, these criteria were, in fact, identical. This way, inter-rater reliability can be computed. Merely calculating a total count could produce the illusion of inter-rater agreement since a frequency count of 3, as in the example, can be achieved in multiple ways.

Inter-rater reliability of the two coders for the present/absent coding

First, inter-rater reliability of the two independent coders for the present/absent SCAN evaluations in the written statements was calculated. As can be seen in Table 2 below, percentages of 1's of coder 1 and coder 2 often deviate considerably from 50%. This indicates a very skewed data set, which in turn leads to the underestimation of Kappa and Phi. Therefore, we decided to primarily use proportion agreement. Table 2 gives a detailed overview of proportion agreement, Phi and Kappa. Proportion agreement was calculated by dividing the number of SCAN evaluations where both coders agreed on the presence or absence of the criterion by the total number of statements. For example, for the *social introduction* criterion, both coders agreed on its presence in 81 out of 82 SCAN evaluations. This results in an inter-rater reliability of $81/82=0.98$. Average proportion agreement for the written statements varied from 0.78 to 1 ($M=0.93$, $SD=0.06$).

11 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Inter-rater reliability of the two independent coders for the recommendations was calculated in the same way, and varied between 0.81 and 1 ($M=0.90$, $SD=0.05$; see Table 2). Kappa and Phi were also calculated and are presented in Table 2. For both the written statements and the recommendations, the reliability showed to be sufficient to combine the scores for each criterion between coders. A criterion was coded as present when both coders coded the criterion as present and a criterion was coded as absent when one, or both coders coded the criterion as absent.

-Please insert Table 2 about here –

Inter-rater reliability of the two coders for the frequency coding

As an index of inter-rater reliability for the frequency coding, Spearman rho was calculated between the frequency counts of the two independent coders. Separate analyses were carried out for SCAN analyst 1 and SCAN analyst 2. As the frequencies were very skewed we also included a detailed overview of Ms, SDs, median and skewness for both coders for each analyst (see Table 3). For SCAN analyst 1, the Spearman rho of the two independent coders for each separate criterion varied from 0.90 to 1, with an average of 0.99 ($SD = 0.03$). Spearman rho of the two independent coders for SCAN coder 2 varied from 0.70 to 1, with an average of 0.98 ($SD = 0.08$)². More in-depth analyses of the percentages (see Table 2, McNemar tests), scored by each of the coders showed some significant differences between both coders. Most significant differences were found for criteria described in the recommendations. This is probably due the unstandardized nature of a recommendation report. Each report follows another structure or outline, depending on the SCAN analyst and type of statement. That is, the criteria are not discussed item-by-item in a specific order, but in a rather random continuous text. Despite these

12 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

few differences, inter-rater reliability showed to be sufficient to average the scores between coders.

-Please insert Table 3 about here –

In sum, we will first investigate the presence of the different criteria within the written statements and the recommendations. Second, we will analyse whether some criteria are more prominent in either the (i) written statements or the (ii) recommendations. Third, we will investigate the inter-rater reliability between the SCAN analysts to see whether SCAN can be used in a reliable way.

Results

Based on the present/absent coding we calculated the proportions of written statements and recommendations in which each SCAN criterion was present, which is presented in Figure 1. For the written statements, results indicated that ten of the criteria were present in more than 25% of the statements (denoted with a * in Figure 1). Furthermore, results indicated that 11 criteria were present in more than 25% of the recommendations (denoted with a * in Figure 1). Nine of these criteria overlapped between the written statements and the recommendations. These criteria were: “Use of pronouns”, “Social introduction”, “Unimportant information”, “Missing information”, “Unasked explanation”, “Negative language”, “Together with”, “Structure of the statement”, and “Change in language”. Three of the criteria differed notably between the written statements and the recommendations. The criterion “Communication” was present in over 25% of the written statements, but not in the recommendations. The criteria “First person singular, past tense” and “Place of emotions” were present in over 25% of the recommendations, but not in the written

13 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

statements. Taking into account all criteria that were present in more than 25% of either the written statements or the recommendations results in a list of 12 SCAN criteria. These criteria are denoted with a * in Figure 1 in either the written statements or the recommendations, or both.

To investigate whether some criteria were more prominent in the written statements or the recommendations, a series of McNemar tests were carried out. To correct for multiple testing, a significance level of 0.01 was used. Only 15 out of the 21 criteria could be scored in the Written Statements, therefore we can only calculate 15 McNemar tests, the remaining six criteria were only discussed in the Recommendations. Nine out of 15 criteria were found to be significantly more often present in the written statements than in the recommendations (see Figure 1). These criteria were “Use of pronouns” [$\chi^2(1, N=82) = 36.03, p<0.01$], “Unimportant information” [$\chi^2(1, N=82) = 28.20, p<0.01$], “Together with” [$\chi^2(1, N=82) = 9.03, p<0.01$], “Missing information” [$\chi^2(1, N=82) = 32.24, p<0.01$], “Unasked explanation” [$\chi^2(1, N=82) = 33.58, p<0.01$], “Negative language” [$\chi^2(1, N=82) = 18.23, p<0.01$], “Communication” [$\chi^2(1, N=82) = 49.02, p<0.01$], “Place of emotion” [$\chi^2(1, N=82) = \text{n.c.}, p<0.01$] and “Social introduction” [$\chi^2(1, N=82) = \text{n.c.}^1, p<0.01$]. None of the remaining criteria were more often present in the Written Statements compared to the Recommendations. As a result, no differences were found for “Verb leaving”, “Structure of the statement”, “Change in language”, “Order”, “Activities”, and “Resistance”.

-Please insert Figure 1 about here –

14 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Next, the inter-rater reliability of the two SCAN analysts was calculated by means of proportion agreement. As the presence of each criterion was coded per line, we were able to compare the frequency counts of both analysts with great precision. An example will help to clarify how reliability was calculated. Consider SCAN analyst 1 who highlights 3 pronouns in line 1, 2 pronouns in line 2, and 5 pronouns in line 3. SCAN analyst 2, in contrast, highlights 4 pronouns in line 1, 2 pronouns in line 2, and 4 pronouns in line 3. First, we look at the comparison between both analysts for line 1. Here we see that they agree about 3 pronouns. For line 2, they agree about 2 pronouns and for line 3 they agree about 4 pronouns. Statements were also compared to make sure that when both analysts coded the same number of criteria within one line; the criteria were, in fact, identical. This leads to a total amount of 9 pronouns on which both analysts agree. However, in total there are 11 (4+2+5) pronouns scored in the three lines. This leads to a proportion agreement of $9/11 = .81$. This calculation was carried out for each criterion in each statement. Next, we calculated the average proportion agreement per criterion for all 82 statements (See Table 4). Average proportion agreement is 0.31 ($SD = 0.21$). The frequency counts of each criterion were averaged for both analysts to obtain the total frequency count of the different criteria within the written statements, which is presented in Table 5.

- Please insert Tables 4 and 5 about here -

Finally, we determined these frequencies for victims ($n = 43$), suspects ($n = 18$), and witnesses ($n = 21$), separately. The total frequencies for these categories were analysed by one-way Analyses of Covariance to evaluate differences between the writers in the presence of criteria, with length of the statement as a covariate. To

15 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

correct for multiple testing, a more conservative significance level of 0.01 was used. However, no significant differences for any of the criteria were found between victims, suspects, and witnesses. Table 5 presents the frequencies of the different criteria for each role per 50 lines, the average length of a statement.

Discussion

Given the unstandardized nature of the SCAN method, the present study examined which SCAN criteria are most frequently used in the field. To address this issue, 82 cases on which police AMS performed SCAN were investigated. Ten criteria were found to be present in over 25% of the written statements and 11 criteria were found to be present in over 25% of the recommendations. Taken together, this resulted in a list of 12 unique criteria that were used in more than 25% of the cases.

These 12 criteria largely overlap with the criteria reported by Smith (2001), and also by Vrij (2008b) in his review of SCAN. The only criterion that is not listed by these authors is the “together with” criterion, which contributed to the recommendations for follow-up investigations in 40% of the cases. Two other criteria (“unasked explanation” and “unimportant information”) are not directly discussed in Vrij (2008a), but these criteria show some degree of overlap with Vrij’s descriptions of the criteria “out of sequence information” and “extraneous information”.

Our results also show that 8 criteria were more often present in written statements than in the recommendations, regardless of the examinee’s status (i.e., victim, suspect or witness). Two criteria were more often present in the recommendations than in the written statements. These criteria were “First person singular, past tense” and “First sentence”. This can be explained by the absence of a coding symbol with which their presence can be marked in written statements. Even

16 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

though “First person singular, past tense” was not scored in the written statements, it emerged in follow-up decisions in almost 60% of the cases (Figure 1).

To evaluate the criteria used in the written statements and the reliability of the SCAN method, we calculated the exact frequency of each criterion. Although the SCAN analysts were trained in using SCAN, their inter-rater reliability was found to be disappointingly low. This finding is especially striking, as all 4 analysts employed the same definitions – given in a self produced summary - and should therefore have been able to investigate these criteria and their appropriate interpretations in a similar way. This low inter-rater reliability suggests that the SCAN method is insufficiently developed as a forensic tool.

A number of the most commonly cited SCAN criteria overlap with criteria from other verbal veracity assessment methods. The criterion “place of emotion”, for example, overlaps with the criterion “accounts of subjective mental state” of the Criteria Based Content Analysis (CBCA; Steller & Köhnken, 1989), and the criterion “affect” of the Reality Monitoring approach (RM; Johnson & Raye, 1981; Sporer, 1997). However, in SCAN, it is primarily the place of the emotion in the statement that is of importance. More specifically, according to SCAN, truth-tellers include emotions particularly after the climax of the story, while liars tend to mention emotions just before the climax of the story. In contrast, in CBCA and RM the presence of emotional language per se is an indication of authenticity regardless of its precise placement. For example, “I was very scared when he touched me” would fulfil the criterion for CBCA and RM regardless of where this sentence is mentioned within the statement. The SCAN criterion “spontaneous corrections” is similar to CBCA criterion 14, also called spontaneous corrections (for a detailed description of CBCA see Vrij, 2008a), and refers to the presence of corrections within the statement. For

17 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

SCAN, their presence indicates deceit, while CBCA analysts believe that their presence indicates truthfulness. A similar discrepancy occurs for the SCAN criterion “lack of conviction or memory”, which comes close to CBCA criterion 15 with the same name. CBCA analysts believe that the presence of this criterion indicates truthfulness, while SCAN users hold the opposite belief. The criteria “out of sequence information” and “extraneous information” are similar to the CBCA criteria “unstructured production” and “superfluous details”, but again, both approaches attach opposite qualities to the presence of these criteria. Research has shown that CBCA criteria are significantly more present in truthful statements relative to false statements (Vrij, 2005). Things are quite different for the related SCAN criteria: recent research has found no evidence that these SCAN criteria are more present in false statements than in truthful statements (Nahari, Vrij, & Fisher, 2011).

Support for certain SCAN criteria comes from research on linguistics features of false and true statements. Newman, Pennebaker, Berry, and Richard (2003) reported that compared to truth-tellers, liars use fewer self-references (e.g., I, me and my) in their stories. These findings were supported by DePaulo and colleagues (2003). By using fewer self-references, liars tend to distance themselves from their statements. The SCAN criteria “first person singular, past tense” and “pronouns” are related to this finding. Sapir (2005) contends that the presence of these two criteria indicates deception, which is in accordance with the findings of the Depaulo et al. (2003) and Newman et al. (2003).

Two limitations of the present dataset deserve attention. First, all the data generated by the vice squad of police AMS were sexual abuse cases. It is possible that some or even all of the most frequently used criteria are specific for these cases. However, the available literature about SCAN describes similar criteria (Vrij, 2008a;

18 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Smith, 2001), which indicates that these criteria are probably also common for non-vice cases. Second, the pilot only included four SCAN trained analysts. Given the unstandardized nature of SCAN, it is possible that these results do not generalize to other SCAN users. Still, the list of most frequently used criteria largely overlaps with those reported in the literature (Vrij, 2008a; Smith, 2001), supporting the notion that these criteria indeed drive SCAN.

In sum, the results of the present study show that SCAN is largely driven by a set of 12 criteria. However, it is important to note that the inter-rater reliability of those criteria is low, except for “Pronouns” and “Social introduction”. As sufficient inter-rater reliability is one of the requirements for any tool to be applied in practice, these findings suggest that SCAN is insufficiently developed as a forensic tool and the extensive use of SCAN in practice should be discouraged. Still, to our knowledge, our study is the first to actually show which criteria drive SCAN. We therefore recommend using the 12 SCAN criteria as a starting point for future studies on their psychometric properties. To overcome the issue of low inter-rater reliability future studies could use a more standardized coding system for scoring SCAN, as was shown by Nahari et al. (2011).

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.
- Akehurst, L., Kohken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology, 10*, 461-471.
- Bockstaele, M. (2008a). De SCAN als middel tot waarheidsvinding [SCAN as a tool for searching the truth]. *Het Tijdschrift voor de Politie [Journal of Police], 70*, 8-13.
- Bockstaele, M. (2008b). Scientific Content Analysis (SCAN). Een nuttig instrument bij verhoren? [SCAN: A valuable tool for interrogations?]. In L. Smets & A. Vrij (Eds.), *Het analyseren van de geloofwaardigheid van verhoren: Het gebruik van leugendeteciemethoden [The analysis of the credibility of interrogations: The use of lie detection methods]* (pp. 105-156). Brussels, Belgium: Politeia.
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Individual Differences, 10*, 214-234.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74-118.
- Driscoll, L. (1994). A validity assessment of written statements from suspects in criminal investigations using the SCAN technique. *Police Studies, 4*, 77-88.
- Greer, E. (2000). The truth behind legal dominance feminism's "two percent false rape claim" figure. *Loyola of Los Angeles Law Review, 33*, 947-972.
- Gudjonsson, G. H., Sigurdsson, J. F., Asgeirsdottir, B. B., & Sigfusdottir, I. D. (2007). Custodial interrogation: What are the background factors associated

20 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

with claims of false confession to police? *Journal of Forensic Psychiatry & Psychology*, *18*, 266-275.

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

Lisak, D., Gardinier, L., Nicksa, S. C., & Cote, A. M. (2010). False allegations of sexual assault: An analysis of ten years of reported cases. *Violence Against Women*, *16*, 1318-1334.

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.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistics styles. *Personality and Social Psychology Bulletin*, *29*, 665-675.

Sapir, A. (2005). *The LSI course on scientific content analysis (SCAN)*. Phoenix, AZ: Laboratory for Scientific Interrogation.

Smith, N. (2001). Reading between the lines: an evaluation of the Scientific content Analysis technique (SCAN). *Police Research Series Paper 135*, 1-42.

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

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

21 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

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. (2008a). Beliefs about nonverbal and verbal cues to deception. In A. Vrij (Ed.), *Detecting lies and deceit* (pp. 115-140). Chichester: Wiley.

Vrij, A. (2008b). Nonverbal dominance versus verbal accuracy in lie detection: A plea to change police practice. *Criminal Justice and Behavior*, *35*, 1323-1336.

Vrij, A., Granhag, P. A., & Porter, S. B. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, *11*, 89-121.

22 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Appendix A

Description of SCAN criteria

The SCAN manual does not consist of a clear list of criteria and definition. Instead all criteria, explanations and examples are spread throughout the entire manual. Therefore we refer to the definitions given by police AMS who studied the manual in great detail and extracted the information from the manual.

1. *Social introduction**: This criterion refers to whether persons that are introduced in the statement are introduced by their name and role (e.g. My son, David). When a person is incompletely introduced this could point to a bad relationship between the writer and the introduced person, especially when other persons are introduced correctly (Police AMS; Sapir, 2005, point 236-258, p. 69-73)
2. *First person singular, past tense**: This criterion is also called the test of commitment which states that a truthful person will write his/her statement in the first person singular, past tense. Especially deviations from past tense or writing in third person could indicate a lack of commitment, which, in turn, could indicate deception (Police AMS; Sapir, 2005, point 179-194, p.47-50).
3. *Unimportant information**: This criterion refers to information that has no function in the statement. This means that the statement could be logically understood without this information. The writer did not have to include this information in the statement but did it anyway. Therefore, according to SCAN, this information is very sensitive and important (Police AMS; Sapir, 2005, point 162d, p. 35)
4. *Use of pronouns**: This criterion refers to the use of pronouns in the statement (e.g., “he”, “my”, “your” etc.). When pronouns are missing within a statement, or more pronouns are expected, this could suggest that the writer wants to distance him/herself from the statement. This could indicate deception (Police AMS; Sapir, 2005, point 260-269, p. 74-77).
5. *Structure of the statement**: In a truthful statement it is expected that 20% is used to write the prologue (e.g., what happened before the main event), 50% is used to describe the main event, and 30% is used to write the epilogue (e.g., discussion about what happened after the event). Large deviations from this structure could indicate deception (Police AMS; Sapir, 2005, point 444-454, p. 119-123).

23 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

6. *Missing information**: This criterion refers to information that is missing and are usually designated by words such as “after a while”, “shortly thereafter”. “the next thing I remember”, etc. It is normal that a person does not tell everything, but missing information situated around the main event could suggest the writer is deliberately hiding important information (Police AMS; Sapir, 2005, point 394-404, p. 110-112).
7. *Out of sequence information**: This criterion refers to information that is given by the writer and has no apparent meaning for the reader. This information feels out of place in the statement but still is regarded as important information since the writer is giving this information for a reason (Police AMS; point 458-459, p. 124-125).
8. *Place of emotions**: For SCAN it is especially the place of the emotions that is of great importance. Usually emotions will be found in the epilogue of the statement. When emotions are already described during the main event, it is possible that the writer is deceptive (Police AMS)
9. *Change in language**: This criterion refers to a change of terminology or vocabulary in the statement. Especially important are words related to categories such as family members, people, communication, transport or weapons. When a change of language is obvious in a statement (e.g. weapon to gun) and no justification can be found for such a change, this could indicate deception (Police AMS; Sapir, 2005, point 162f-162j, p. 37-38 and point 271, p. 80, and point 461-473, p. 125-127).
10. *Resistance during rape**: With SCAN it is expected that victims of sexual abuse write something about how they tried to resist the offence. When there is no resistance mentioned in the statement this may indicate deception (Police AMS).
11. *First sentence**: According to SCAN the first sentence is a very important sentence in the statement. A lot of information can be found in de first sentence (Police AMS; Sapir, 2005, point 123, p. 27).
12. *Order**: This criterion refers to the order in which persons or objects are mentioned in the statement. In this way the writer reveals his/her priority regarding these persons or objects (Police AMS; Sapir, 2005, point 259, p. 73).

24 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

13. *Verb leaving**: According to SCAN the verb leaving is important. Using this term in the statement may indicate deception, especially when this verb is used in the first sentence (Police AMS; Sapir, 2005, point 273-275, p. 80-81).
14. *Communication**: According to SCAN every verb in relation to communication is important. When a writer is able to cite parts of conversations in the statement this indicates truthfulness (Police AMS; Sapir, 2005, point 312, p. 91).
15. *Objective versus subjective time**: This criterion refers to the relationship between subjective and objective time. Subjective time is the pace of the statement; more precisely how many lines are necessary to write about one objective hour. On average it is expected that a writer needs 3 or 4 lines per hour when describing one day. Large deviations from this pace suggest deception (Police AMS, Sapir, 2005, point 280-295, p. 81-86).
16. *Extraneous information**: This criterion refers to information that does not seem relevant or seems strange for the reader. According to SCAN the use of this information may be used to divert attention from other information which can indicate deception (Police AMS, Sapir, 2005, point 173a-173b, p. 42-43)
17. *Together with**: According to SCAN the use of the pronoun “we” indicate that the writer feels a certain commitment to the other person. However, when a writer uses the term “together with” there is a lower sense of commitment to the other person. This information is used to highlight tension between the different persons mentioned in the statement (Police AMS; Sapir, 2005, point 270, p. 78-79).
18. *Details*: This criterion refers to the usage of details in the statement. If the writer is able to produce a lot of details in the statement, the statement is expected to be truthful (Bockstaele, 2008a, 2008b).
19. *Unasked explanation**: This criterion refers to an explanation why something happened, given by the writer, without asking. According to SCAN this information is very sensitive (Police AMS; Sapir, 2005, point 299a, p. 86-87).
20. *Lack of conviction or memory*: This criterion refers to vagueness in the statement about certain event (e.g. “I think...”, “I suppose...”) or parts in the statement where the writer states that he or she cannot remember something. These cues are seen as an indication of deceit (Vrij, 2008b).

25 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

21. *Unexpected complication*: This criterion refers to the presence of unexpected events. For example, a rape victim who states that her attacker first had to get the cat off the bed. It is expected that the presence of this criterion indicates truthfulness (Bockstaele, 2008a, 2008b).
22. *Denial of allegation*: When the writer directly denies the allegation in the statement it indicates truthfulness (e.g. "I did not..."). For deceptive persons it is expected they will not directly deny the allegation (e.g. "Do you think I would do something like that?") (Vrij, 2008b; Sapir, 2005, point 206, p. 57).
23. *Spontaneous corrections*: This criterion refers to the presence of corrections in the statement. Before examinees write their statement they are instructed not to cross anything out. When examinees do not follow this instruction it may indicate deception (Vrij, 2008b).
24. *Sensory perceptions*: The use of sensory perceptions within a statement is an indication of truthfulness (Bockstaele, 2008a, 2008b).
25. *Unresponsiveness to questions*: When a writer refuses to write down information this is an indication of deception (Bockstaele, 2008b).
26. *Avoiding answers*: The unwillingness of examinees to give direct answers to questions indicates deception (Bockstaele, 2008b).
27. *Activities**: According to SCAN certain discussed activities are important. These activities include brushing teeth, turning the light on or off, closing or opening a door or getting in or out of a car. These activities can give information about deception or child sexual abuse (Police AMS; Sapir, 2005, point 303, p. 89).
28. *Exact location**: When a writer gives an exact location of another person in the statement this gives an indication about a conflict between the writer and the other person (Police AMS).
29. *Negative language use**: When a writer gives information about something that did not happen, thus when a sentence is presented in negative. This is sensitive information for the writer. (Police AMS; Sapir, 2005, point 299c, p. 87). This criterion is a combination of "Denial of allegation" and "Lack of conviction or memory".

26 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Table 1.

Literature overview of used SCAN criteria and data police AMS)

References	Bockstaele (2008a)	Bockstaele (2008b)	Driscoll (1994)	Smith (2001)	Nahari et al (2011)	Police AMS
1. Social introduction #	x	x	-	x	x	x (coded green)
2. First person singular, past tense <	x	x	x	x	x	x (no code)
3. Unimportant information #	-	-	-	-	-	x (coded yellow)
4. Use of pronouns <	x	x	x	x	x	x (circled red)
5. Structure of the statement <	x	x	x	x	x	x (marked in margin)
6. Missing information #	x	x	x	x	x	x (coded pink)
7. Out of sequence information #	x	x	x	x	x	x (no code)
8. Place of emotions <	x	x	x	x	x	x (underlined)
9. Change in language <	x	x	x	x	x	x (words are linked)
10. Resistance during rape <	-	-	-	-	-	x (marked in margin)
11. First sentence #	-	-	-	-	-	x (no code)
12. Order #	-	-	-	-	-	x (marked in margin)
13. Verb leaving <	-	-	-	-	-	x (coded blue)
14. Communication >	-	-	-	-	-	x (coded orange)
15. Objective vs. subjective time <	x	x	-	x	x	x (no code)
16. Extraneous information <	-	-	-	-	x	x (no code)
17. Together with #	-	-	-	-	-	x (coded purple)
18. Details >	x	x	-	-	-	-
19. Unasked explanation #	-	-	-	-	-	x (coded blue)
20. Lack of conviction or memory <	x	x	x	x	x	-
21. Unexpected complications >	x	x	-	-	-	-
22. Denial of allegation >	-	-	x	x	x	-
23. Spontaneous corrections <	x	x	x	x	x	-

27 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

24. Sensory perceptions >	x	x	-	-	-	-
25. Unresponsiveness to questions <	-	x	-	-	-	-
26. Avoiding answers <	-	x	-	-	-	-
27. Activities <	-	-	-	-	-	x (underlined)
28. Exact location #	-	-	-	-	-	x (no code)
29. Negative language #	-	-	-	-	-	x (coded yellow)
Total	14	16	10	12	13	21

Note. Criteria described by the authors are indicated with an “x”. Criteria marked with ‘<’ indicate deception, criteria marked with ‘>’ indicate truthfulness. When indicated with ‘#’ see appendix A for further information. For the coding information of police AMS see The Amsterdam Police Pilot on p.5-6.

28 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Table 2.

Detailed overview of inter-rater reliability of the two independent coders who coded the presence/absence of the criteria in one or both of the written statements(top panel) and the recommendations (bottom panel) produced by the SCAN analysts.

Criteria	Written statements					
	Kappa	Phi	Percentage Agreement	Percentage present coder 1	Percentage present coder 2	McNemar
4. Pronouns	1	1	1	100	100	nc
19. Unasked explanation	0.73	0.73	93.98	83.13	85.54	ns
13. Verb leaving	0.60	0.61	89.16	18.07	14.46	ns
3. Unimportant info	0.26*	0.26*	93.98	95.18	96.39	ns
14. Communication	0.89	0.89	95.18	65.06	67.47	ns
1. Social introduction	nc	nc	98.80	100	98.80	nc
6. Missing info	1	1	1	89.16	89.16	ns
5. Structure	0.76	0.76	87.95	51.81	51.81	ns
9. Change in language	0.87	0.87	93.98	36.14	32.53	ns
29. Negative language	0.39*	0.40*	78.31	71.08	79.52	ns
17. Together with	0.81	0.81	91.57	68.67	65.06	ns
27. Activities	1	1	1	0	0	nc
12. Order	0.21	0.22	92.77	3.61	6.02	ns
8. Emotion	0.63	0.66	89.16	31.33	19.28	0.006

29 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

10. Resistance	0.77	0.77	92.77	20.48	18.07	ns
Recommendations						
4. Pronouns	0.78	0.79	92.77	59.04	53.01	ns
7. Out of sequence info	0.52	0.59	93.98	9.64	3.61	ns
19. Unasked explanation	0.85	0.86	93.98	40.96	33.73	ns
13. Verb leaving	0.63	0.68	92.77	14.46	7.23	ns
3. Unimportant info	0.78	0.80	87.95	61.45	49.40	0.004
14. Communication	0.42	0.52	93.98	8.43	2.41	ns
1. Social introduction	0.86	0.86	91.57	72.29	68.67	ns
6. Missing info	0.83	0.84	91.57	48.19	39.76	ns
2. Commitment	0.69	0.72	86.75	74.70	63.86	ns
5. Structure	0.74	0.77	87.95	69.88	57.83	0.002
15. Obj vs subj time	0.06 (ns)	0.09 (ns)	81.93	16.87	3.614	0.007
9. Change in language	0.88	0.89	91.57	36.14	26.51	ns
16. Extraneous info	nc	nc	86.75	13.25	0	nc
11. First sentence	0.55	0.56	81.93	38.55	19.28	<0.001
28. Exact location	0.50	0.50	91.57	13.25	4.82	ns
29. Negative language	0.92	0.92	96.39	32.53	28.92	ns
17. Together with	0.65	0.69	81.93	57.83	39.76	<0.001
27. Activities	1	1	1	0	0	nc

30 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

12. Order	0.44	0.53	86.75	21.69	7.23	<0.001
8. Emotion	0.65	0.69	81.93	57.83	39.76	<0.001
10. Resistance	0.63	0.68	86.75	28.92	15.66	0.001

Note. All p 's < 0.001, p *'s < 0.05. ns indicates not significant. nc indicates that these results could not be calculated because at least the score of 1 coder was seen as a constant (all 1 or all 0 scores). Kappa and Phi are often very low due to the skewness of the data, nevertheless percentage agreement is very high. Numbers of criteria refer to the description of the particular criterion in Appendix A.

31 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Table 3.

Detailed overview of the frequency coding of the two independent coders for the written statements, separated for each SCAN analyst

Criteria	Analyst 1									
	Coder 1				Coder 2				Coder 1 and 2	
	M	SD	Median	Skewness	M	SD	Median	Skewness	Spearman rho	
4. Pronouns	53.89	39.12	50	1.022	53.9	39.11	50	1.02	1	
19. Unasked explanation	2.72	3.08	2	1.65	2.72	3.08	2	1.65	1	
13. Verb leaving	0.56	1.07	0	3.09	0.56	1.07	0	3.09	1	
3. Unimportant information	3.04	3.2	2	1.32	3.05	3.22	2	1.32	1	
14. Communication	3.21	4.15	2	1.5	3.21	4.15	2	1.5	1	
1. Social introduction	15.84	14.47	11	1.65	15.86	14.45	11	1.65	1	
6. Missing info	3.77	4.26	2	1.51	3.77	4.26	2	1.51	1	
5. Structure	0.42	0.52	0	0.6	0.42	0.52	0	0.6	1	
9. Change in language	0.25	0.49	0	1.84	0.25	0.49	0	1.86	1	
29. Negative language	2.01	2.56	1	2.3	2.02	2.55	1	2.31	0.99	
17. Together with	1.64	2.42	0	1.67	1.53	2.29	0	1.83	0.9	
27. Activities	0	0	0	0	0	0	0	0	nc	
12. Order	0.01	0.11	0	9	0.01	0.11	0	9	1	
8. Emotion	0.53	2.49	0	8.27	0.53	2.49	0	8.27	1	
10. Resistance	0.14	0.44	0	4.28	0.14	0.44	0	4.25	1	

32 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Criteria	Analyst 2								
	Coder 1				Coder 2				Coder 1 and 2
	M	SD	Median	Skewness	M	SD	Median	Skewness	Spearman rho
4. Pronouns	45.51	44.4	34	1.38	45.51	44.4	34	1.38	1
19. Unasked explanation	2.18	2.67	1	1.5	2.21	2.72	1	1.51	1
13. Verb leaving	0.45	0.8	0	2.21	0.45	0.8	0	2.21	1
3. Unimportant information	2.57	3.12	2	1.84	2.57	3.12	2	1.84	1
14. Communication	3.32	4.97	1	2.45	3.32	4.97	1	2.45	1
1. Social introduction	14.37	17.82	8	1.99	14.39	17.84	8	1.99	1
6. Missing info	3.59	4.51	2	1.7	3.66	4.62	2	1.79	1
5. Structure	0.41	0.57	0	1.39	0.41	0.57	0	1.39	1
9. Change in language	0.28	0.5	0	1.57	0.28	0.5	0	1.57	1
29. Negative language	1.65	2.36	1	2.62	1.73	2.41	1	2.43	0.99
17. Together with	1.71	2.94	0	2.31	1.71	2.93	0	2.33	0.99
27. Activities	0	0	0	0	0	0	0	0	nc
12. Order	0.01	0.11	0	9.06	0.02	0.16	0	6.28	0.7
8. Emotion	0.39	2.04	0	8.16	0.39	2.04	0	8.16	1
10. Resistance	0.15	0.52	0	4.39	0.15	0.52	0	4.39	1

Note: Numbers of criteria refer to the description of the particular criterion in Appendix A. nc indicates Spearman rho was not calculated because the variables were constant.

33 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Table 4.

Detailed overview of inter-rater reliability of the SCAN analysts for the written statements

Criteria	Proportion agreement
4. Pronouns	0.79
19. Unasked explanation	0.29
13. Verb leaving	0.32
3. Unimportant information	0.18
14. Communication	0.27
1. Social introduction	0.65
6. Missing info	0.30
5. Structure	0.32
9. Change in language	0.14
29. Negative language	0.30
17. Together with	0.46
27. Activities	0
12. Order	0
8. Emotion	0.49
10. Resistance	0.2
Average	0.31

Note. Only 15 criteria were coded within the statements so inter-rater reliability could only be calculated for these 15. Numbers of criteria refer to the description of the particular criterion in Appendix A.

34 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

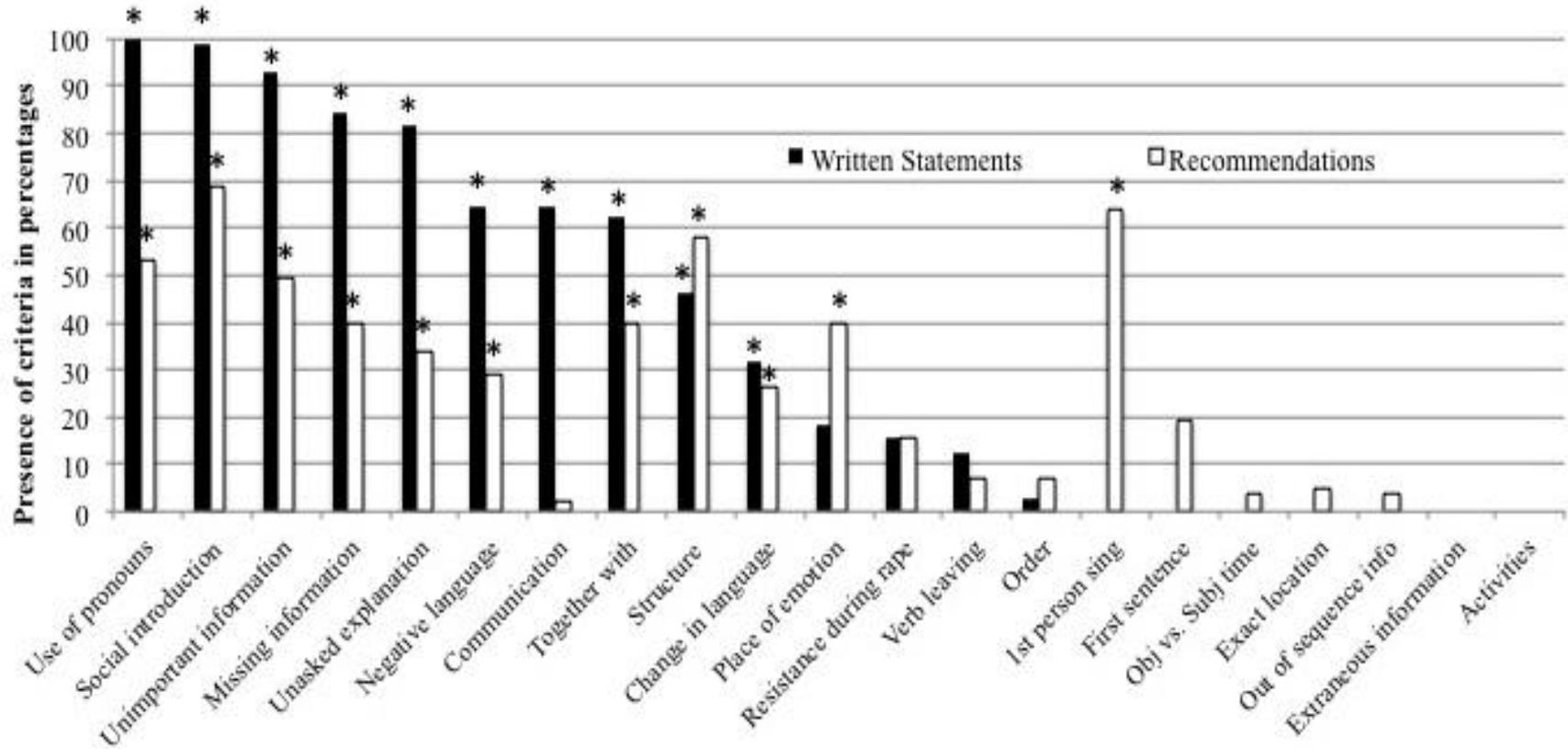
Table 5.

Total frequency count for each criterion and frequency separated for the role of the writer, corrected for the length of the statement

Criteria	Total Frequency	Mean (SD)	Median	Interquartile range	Frequency victims 50 lines	Frequency suspects 50 lines	Frequency witness 50 lines
4. Use of pronouns	4463.50	54.43 (38.79)	52	49.75	55.15	47.29	43.53
1. Social Introduction	1320	16.10 (14.25)	10.5	16.75	12.13	20.39	17.87
6. Missing Information	319	3.89 (3.75)	3	5.38	4.07	4.57	2.11
14. Communication	286	3.49 (4.43)	2	4.50	3.19	3.27	3.38
3. Unimportant information	247.5	3.02 (2.90)	2	3.50	3.01	2.32	2.70
19. Unmasked explanation	232	2.83 (2.75)	2	3.88	2.75	3.62	1.81
29. Negative language	167.50	2.04 (2.29)	1.5	2	2.15	0.74	0.89
17. Together with	146	1.78 (2.26)	1	2.88	1.50	2.53	1.48
13. Verb leaving	41.50	0.51 (0.81)	0	1	0.52	0.46	0.38
8. Place of emotion	40	0.49 (2.24)	0	0.50	0.70	0.25	0.06
5. Structure	36.50	0.45 (0.43)	0.5	0.5	0.44	0.50	0.31
9. Change in language	23.50	0.29 (0.41)	0	0.5	0.27	0.39	0.19
10. Resistance	12	0.15 (0.43)	0	0	0.23	0	0.02
12. Order	1.50	0.02 (0.10)	0	0	0	0.11	0
27. Activities	0	0 (0)	0	0	0	0	0

Numbers of criteria refer to the description of the particular criterion in Appendix A.

35 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA



36 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Figure 1. Overview of the SCAN criteria that were present in the written statements, and the criteria that were present in the recommendations for further investigation. Criteria marked with a * were present in more than 25% of the written statements or the recommendations. The X-axis gives the name of each criterion, definitions can be found in Appendix A

37 STATEMENTS REVEAL SCAN IS LARGELY DRIVEN BY 12 CRITERIA

Notes

1. No McNemar statistic could be computed, because one of the variables showed no variation.
2. High correlations were expected as differences between coders merely reflected counting errors