One approach to detecting lies that has a long history is to analyse speech content. In around 900 BC, a papyrus of the Vedas described how to identify a poisoner. In addition to some physiological and behavioural cues, it says that a poisoner ‘does not answer questions, or they give evasive answers; he speaks nonsense’ (Trovillo, 1939, p. 849). Tardieu, a French forensic expert, already acknowledged in the 1850s that in children’s alleged sexual abuse cases, certain characteristics in the story needs to be considered, such as ‘quantity of detail’ (Lamers-Winkelman, 1999). Walker (1886), an American forensic medical doctor, claimed that mere reliance upon the physical examination in alleged child sexual abuse cases is unreliable. Rather, according to Walker, children should be encouraged to tell their stories in their own words, and the way in which children tell their stories and the expressions they
use are amongst the best guides to distinguish truth from deception in children (see Lamers-Winkelman, 1999).

Since the 1950s, the search for verbal cues to deceit has accelerated, and to date, a substantial number have been found to be diagnostic for deceit (DePaulo et al., 2003; Masip, Sporer, Garrido, & Herrero, 2005; Vrij, 2008). Some verbal cues are measured in isolation, but most are examined as part of verbal veracity assessment tools and this chapter discusses three of those tools: Statement Validity Analysis (SVA), Reality Monitoring (RM) and Scientific Content Analysis (SCAN). The focus is on those three tools because they are frequently used by practitioners and/or scholars. SVA is the most frequently researched verbal veracity tool to date and more than 50 studies have been published examining the working of this tool. It is also frequently used in daily life as SVA assessments are used as evidence in criminal courts in several European countries. RM is, to our knowledge, never used in real life but it is popular amongst scholars, perhaps because it has a solid theoretical background. The RM deception research body is substantial and more than 30 RM deception studies have been carried out to date. Conversely, SCAN is very popular in the field but has hardly been researched. This chapter provides only outlines of the three tools. For more detailed information about SVA, see Gumpert and Lindblad’s (1999), Köhnken (2004), Raskin and Esplin (1991), Steller and Boychuk (1992) and Vrij (2005, 2008). For more detailed information about RM, see Masip et al. (2005), Sporer (2004) and Vrij (2008). For more detailed information about SCAN, see Armistead (2011), Driscoll (1994), Nahari, Vrij, and Fisher (2012) and Smith (2001).

The final section of this chapter compares the three tools by using the set of guidelines provided by the United States Supreme Court for admitting expert scientific evidence in (American) federal courts. These guidelines give good opportunity to summarize the key aspects of the tools, the extent to which they have been examined and the empirical and academic support each.

More than three verbal tools are available in the deception literature, of which two deserve special mention because they have potential. The Assessment Criteria Indicative of Deception (ACID) combines interview techniques aimed at eliciting cues to deception with the analysis of empirically derived content criteria (see Colwell, Hiscock-Anisman, and Fede (2013) for a review). The Aberdeen Report Judgment Scales are based on a factor analysis of RM and CBCA criteria and theoretical considerations such as autobiographical memory, the social psychology of attribution, and impression management and deception (see Sporer (2004, 2012) for overviews).
STATEMENT VALIDITY ANALYSIS

Statement Validity Analysis is a tool designed to determine the credibility of child witnesses’ testimonies in trials for sexual offences. It is not surprising that a technique has been developed to verify whether or not a child has been sexually abused. It is often difficult to determine the facts in an allegation of sexual abuse, since often there is no medical or physical evidence. Frequently, the alleged victim and the defendant give contradictory testimony, and often there are no independent witnesses to give an objective version of events. This makes the perceived credibility of the defendant and alleged victim important. The alleged victim is in a disadvantageous position if he or she is a child, as adults have a tendency to mistrust statements made by children.

SVA assessments are accepted as evidence in some North American courts and in criminal courts in several Western European countries including Germany, the Netherlands, Spain and Sweden (Vrij, 2008). The tool originates from Sweden (Trankell, 1972) and Germany (Undeutsch, 1982) and consists of four stages (Vrij, 2008): (i) a case-file analysis; (ii) a semi-structured interview; (iii) a Criteria-Based Content Analysis (CBCA) that systematically assesses the quality of the transcribed interviews and (iv) an evaluation of the CBCA outcome via a set of questions (Validity Checklist). Much of the SVA research is concerned with the ability of CBCA, one of the four SVA stages, to discriminate between truth tellers and liars. Also the Validity Checklist, another stage of the SVA procedure, has attracted attention from researchers.

A Case-File Analysis

The SVA procedure starts with the analysis of the case file. A case file should include information about the child witness (e.g. his or her age, cognitive abilities, relationship to the accused person), the nature of the event in question and previous statements of the child and other parties involved. The case-file analysis gives the SVA expert insight into what may have happened and the issues that are disputed. The SVA analysis focuses on these disputed elements in the subsequent three stages.

A Semi-Structured Interview

The second stage of SVA is a semi-structured interview where the child provides his or her own account of the allegation. Conducting a proper interview is never an easy task, but interviewing young children is
Detecting Deception particularly difficult, because their descriptions of past events are notably incomplete (Bull, 2010; Goodman & Melinder, 2007). Therefore, interviewers routinely want more information than is initially provided (Kebbel & Milne, 1998), and interviewers have to ask further, specific questions to learn more about an event. The danger that interviewers face is that their questioning may become suggestive. In that case, the question suggests to the child what the answer should be and, subsequently, leads the child to providing that answer.

Special interview techniques based upon psychological principles have been designed to obtain as much information as possible from interviewees in a free narrative style without inappropriate prompts or suggestions. The UK Home Office (2002) and the American Professional Society on the Abuse of Children (1997) as well as many researchers provide guidance for good interviewing techniques (Bull, 2010; Fisher, 2010; Lamb, Hershkowitz, Orbach, & Esplin, 2008), but Raskin and Esplin (1991) deserve special mention as they outlined interview procedures specifically for SVA interviews. They stress the importance of the interviewer becoming as familiar with the case material as possible, and that the interview should take place in a supportive but neutral environment designed to maximize the performance level of the child witness. They further describe the importance of establishing rapport with the child prior to the interview (by asking the child about something pleasant or interesting that the interviewer knows the child has recently experienced) and the importance of motivating the child to be completely truthful during the interview, for example by discussing the differences between truths and lies. During the actual interview, the interviewer should take great care to use language and concepts that are appropriate for the age and cognitive development of the child. In addition, questions should be asked one at a time, and an answer should be obtained prior to asking the next question. Questions and comments that create expectations (criticism or praise) should be avoided.

The interview should start with a free narrative (‘I understand there is a problem in your family and I need you to tell me about it so that I can help’), and the interviewer must be patient when the child starts answering the question (never interrupt the child). A child can be encouraged to say more through questions such as ‘Did anything else happen?’ The next phase includes asking open-ended questions used to elicit additional detail that seems necessary to complete the descriptions already given by the child (‘You said something also happened in the car, tell me about that?’). Direct questioning should only be used if the elicited information requires clarifications (‘So you said he was sitting next to you and then he put his thing inside you. How did that
happen?’ or ‘Did he take your clothes off?’). Probing questions may be used to obtain information to evaluate alternative hypotheses (‘You told me that he did that when you were only two, but you said you don’t remember it. How do you know what happened?’). The interview should always end on a positive note that leaves the child feeling good, for example, by introducing a topic that is pleasant for the child to discuss, such as plans for the upcoming holiday or birthday.

**CRITERIA-BASED CONTENT ANALYSIS**

The interviews are audiotaped and transcribed, and the transcripts are used for the second part of SVA: the CBCA. Trained evaluators judge the presence or absence of 19 criteria (see Table 1.1). CBCA is based on the hypothesis, originally stated by Undeutsch, that a statement derived from memory of an actual experience differs in content and quality from a statement based on invention or fantasy, known as the Undeutsch Hypothesis (Steller, 1989). The presence of each criterion strengthens the hypothesis that the account is based on genuine personal experience. In other words, truthful statements will have more of the elements measured by CBCA than false statements. A theoretical foundation for the Undeutsch Hypothesis was presented by Köhnken (1989, 1996, 2004), who proposed that both cognitive and motivational factors influence CBCA scores.

With regard to cognitive factors, it is assumed that the presence of several criteria (Criteria 1–13) is likely to indicate genuine experiences as they are typically too difficult to fabricate. Therefore, statements which are coherent and consistent (logical structure), whereby the information is not provided in a chronological time sequence (unstructured production) and which contain a significant amount of detail (quantity of detail) are more likely to be true. Regarding details, accounts are more likely to be truthful if they include contextual embeddings (references to time and space: ‘He approached me for the first time in the garden during the summer holidays’), descriptions of interactions (‘The moment my mother came into the room, he stopped smiling’), reproduction of speech (speech in its original form: ‘And then he asked: Is that your coat?’), unexpected complications (elements incorporated in the statement which are somewhat unexpected, e.g. the child mentions that the perpetrator had difficulty with starting the engine of his car), unusual details (details which are uncommon but meaningful, e.g. a witness who describes that the man she met had a stutter) and superfluous details (descriptions which are not
### Table 1.1 The criteria-based content analysis criteria

**General characteristics**

1. **Logical structure**
   - Coherency of the statement in terms of not containing logical inconsistencies or contradictions.

2. **Unstructured production**
   - The presentation of the information in a (non) chronological order.

3. **Quantity of details**
   - The inclusion of specific descriptions of place, time, persons, objects and events.

**Specific contents**

4. **Contextual embedding**
   - Events being placed in time and location, and actions being connected with other daily activities and/or customs.

5. **Descriptions of interactions**
   - Information that interlinks at least the alleged perpetrator and witness.

6. **Reproduction of conversation**
   - Parts of the conversation are reported in original form or if the different speakers are recognizable in the reproduced dialogues.

7. **Unexpected complications during the incident**
   - Elements incorporated in the statement that are somewhat unexpected.

8. **Unusual details**
   - Details of people, objects or events that are unique, unexpected or surprising but meaningful in the context.

9. **Superfluous details**
   - Details in connection with the allegations that are not essential for the accusation.

10. **Accurately reported details misunderstood**
    - Mentioning of details that are beyond the interviewee’s comprehension.

11. **Related external associations**
    - Events are reported that are not actually part of the alleged offence but are merely related to the offence.

12. **Accounts of subjective mental state**
    - Development and change of feelings experienced at the time of the incident. This criterion also includes reports of thoughts.

13. **Attribution of perpetrator’s mental state**
    - Descriptions of the alleged perpetrator’s feelings, thoughts or motives during the incident.

**Motivation-related contents**

14. **Spontaneous corrections**
    - Corrections that are made or information that is added to material previously provided in the statement without having been prompted by the interviewer.
Another criterion that might indicate truthfulness is when a witness speaks of details that are beyond the horizon of his or her comprehension, for example, when he or she describes the adult’s sexual behaviour but attributes it to a sneeze or to pain (accurately reported details misunderstood). Finally, possible indicators of truthfulness are: if the child reports details which are not part of the allegation but are related to it (related external associations, e.g. a witness who describes that the perpetrator talked about the women he had slept with and the differences between them), when the witness describes his or her feelings or thoughts experienced at the time of the incident (accounts of subjective mental state), or describes the perpetrator’s feelings, thoughts or motives during the incident (attribution of perpetrator’s mental state: ‘He was nervous, his hands were shaking’).

Other criteria (Criteria 14–18) are more likely to occur in truthful statements for motivational reasons. A truthful person will not be as concerned with impression management as a deceiver. Compared to truth tellers, deceivers will be more keen to try to construct a report which they believe will make a credible impression on others, and will leave out information which, in their view, will damage their image of being a sincere person (Köhnenk, 1999). As a result, a truthful

### Table 1.1 (Cont’d)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15. Admitting lack of memory</td>
<td>An unprompted interviewee admitting lack of memory by either saying ‘I don’t know’ or ‘I don’t remember’.</td>
</tr>
<tr>
<td>16. Raising doubts about one’s own testimony</td>
<td>Interviewee indicating that part of his or her description sounds odd, implausible, unlikely, etc.</td>
</tr>
<tr>
<td>17. Self-deprecation</td>
<td>Inclusion of personally unfavourable, self-incriminating details.</td>
</tr>
<tr>
<td>18. Pardoning the perpetrator</td>
<td>Failing to blame the perpetrator or excusing his or her behaviour.</td>
</tr>
<tr>
<td><strong>Offence-specific elements</strong></td>
<td>Descriptive element of the crime that are known by professional to be typical for the type of crime under investigation but are counter-intuitive for the general public.</td>
</tr>
</tbody>
</table>

**15. Admitting lack of memory**

An unprompted interviewee admitting lack of memory by either saying ‘I don’t know’ or ‘I don’t remember’.

**16. Raising doubts about one’s own testimony**

Interviewee indicating that part of his or her description sounds odd, implausible, unlikely, etc.

**17. Self-deprecation**

Inclusion of personally unfavourable, self-incriminating details.

**18. Pardoning the perpetrator**

Failing to blame the perpetrator or excusing his or her behaviour.

**Offence-specific elements**

Descriptive element of the crime that are known by professional to be typical for the type of crime under investigation but are counter-intuitive for the general public.

---

Adapted with permission from Vrij (2008). © John Wiley & Sons, Ltd.
statement is more likely to contain information that is inconsistent with the stereotypes of truthfulness. The CBCA list includes five of these so-called ‘contrary-to-truthfulness-stereotype’ criteria (Ruby & Brigham, 1998): spontaneous corrections (corrections made without prompting from the interviewer (‘He wore black trousers, no sorry, they were green’), admitting lack of memory (expressing concern that some parts of the statement might be incorrect: ‘I think’, ‘Maybe’, ‘I am not sure’, etc.), raising doubts about one’s own testimony (anticipated objections against the veracity of one’s own testimony: ‘I know this all sounds really odd’), self-deprecation (mentioning personally unfavourable, self-incriminating details: ‘Obviously it was stupid of me to leave my door wide open because my wallet was clearly visible on my desk’) and pardoning the perpetrator (making excuses for the perpetrator or failing to blame him or her, such as a girl who says she now feels sympathy for the defendant who possibly faces imprisonment).

The final criterion relates to details characteristic of the offense. This criterion is present if a description of events is typical for the type of crime under investigation but is counter-intuitive for the general public. For example, a witness describes feelings that professionals know are typical for victims of incestuous relationships but that seem counter-intuitive and odd to the layman.

THE VALIDITY CHECKLIST

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

The fourth and final phase of the SVA method is to examine whether any of these alternative explanations might have affected the presence
of the CBCA criteria in the transcripts. For this purpose, a checklist, the Validity Checklist, has been compiled, which comprises 11 issues (called ‘external factors’ hereafter) that are thought to possibly affect CBCA scores. By systematically addressing each of the external factors addressed in the Validity Checklist, the evaluator explores and considers alternative interpretations of the CBCA outcomes. Each affirmative response that the evaluator gives to an external factor raises a question about the validity of the CBCA outcome.

Detailed descriptions of the external factors mentioned in the Validity Checklist are provided by Raskin and Esplin (1991), Steller (1989), Steller and Boychuk (1992) and Yuille (1988). Slightly different versions of the Validity Checklist exist (different authors have used somewhat different versions). The Validity Checklist presented below is that published by Steller and colleagues (Steller, 1989; Steller & Boychuk, 1992). SVA evaluators consider the following external factors:

1. Appropriateness of language and knowledge, for example mental capability of the child;
2. Appropriateness of affect shown by the interviewee;
3. Interviewee’s susceptibility to suggestion;
4. Evidence of suggestive, leading or coercive questioning;
5. Overall adequacy of the interview;
6. Motives to report, for example, whether the interviewee’s relationship with the accused, or other people involved, suggests possible motives for a false allegation;
7. Context of the original disclosure or report, for example, whether there are questionable elements in the context of the original disclosure;
8. Pressures to report falsely, such as indications that others suggested, coached, pressured or coerced the interviewee to make a false report;
9. Consistency with the law of nature, whether the described events are unrealistic;
10. Consistency with other statements, whether there are major elements in the statement that are inconsistent or contradicted by another statement made by this interviewee; and
11. Consistency with other evidence, for example, whether there are major elements in the statement that are contradicted by reliable physical evidence or other concrete evidence.

In the fourth stage of the SVA procedure, evaluation of the CBCA outcome, the evaluator systematically addresses each of the external factors mentioned in the Validity Checklist, and explores and considers alternative interpretations of the CBCA outcomes.

SVA: Research

Despite the fact that SVA assessments are used as evidence in court in several countries, it is unclear how accurate these assessments are because no reliable data regarding the accuracy of SVA assessments in real-life cases are currently available. To examine the accuracy of SVA
assessments in real-life cases, it is necessary to know what truly happened in the disputed event. Obtaining this so-called ground truth is difficult because it can only be determined via case facts, such as medical evidence or other evidence, which indisputably links, or does not link, the alleged perpetrator to the crime. Such case facts are often never present in sexual abuse cases. Instead, researchers often use ‘soft’ evidence as ground truth (such as confessions). It is a common problem in deception field studies that researchers do not know what truly happened in the cases that are included in their field study.

Research has been carried out in the form of laboratory studies, but it has mainly been focused on the third phase of SVA: the accuracy of CBCA assessments. Vrij (2008) reviewed the available CBCA literature, and a summary of this review is presented here. In those studies, either children, but more often undergraduate students, told the truth or lied for the sake of the experiment. Such studies have revealed similar results for adults and children. In alignment with the CBCA assumption, many CBCA criteria were more often present in truthful statements than in fabricated reports. Of the individual criteria, Criterion 3, quantity of details, received the most support. The amount of details was calculated in 29 studies (field studies and laboratory studies combined), and in 22 of those studies (76%), truth tellers included significantly more details into their accounts than liars. Moreover, in not a single study did truth tellers include significantly less details into their statements than liars. This is impressive support for Criterion 3. Other criteria that received strong support were Criterion 4, contextual embeddings (truth tellers included more contextual embeddings in 16 out of 26 studies examining this criterion), and Criterion 6, reproduction of conversation (truth tellers included more reproductions of conversations in 15 out of 25 studies examining this criterion). A total CBCA score was reported in 20 studies, and truth tellers obtained higher CBCA scores than liars in 16 out of those 20 studies, whereas in none of the studies did liars obtain a higher CBCA score than truth tellers. This represents strong support for the CBCA total score.

In 19 laboratory studies, CBCA accuracy rates for classifying truth tellers and liars were calculated. The results revealed that 71% of the truths and 71% of the lies were correctly classified by using CBCA assessments (Vrij, 2008), whereas 50% accuracy rate can be expected by chance alone (by tossing a coin). Whether this reflects the accuracy of CBCA assessments in real-life criminal investigations is unknown. Students or children who tell lies and truths in an experiment are different from children who tell truths and lies in criminal investigations, and the accuracy scores therefore do not necessarily reflect the
accuracy scores in criminal investigations. Three early CBCA field studies have reported accuracy rates, and two of them yielded very high accuracy rates (Esplin, Boychuk, & Raskin, 1988; Parker & Brown, 2000). However, the ground truth in those two studies is unknown, and the results are therefore unreliable. Fortunately, two CBCA field studies have been published more recently (both involving child witnesses) in which ground truth was established in a satisfactory manner. In Akehurst, Manton, and Quandte (2011), only a low number of fabricated cases (N = 10) were included in the study, almost certainly because it is difficult to establish ground truth. In that study, two experts rated the transcripts. One expert classified 81% of the fabricated cases correctly, whereas the other rater classified 60% of the fabricated cases correctly. Roma, San Martini, Sabatello, Tatarelli, and Ferracuti (2011) included many cases in their field study (60 true and 49 false cases). Trained CBCA coders examined the presence or absence of 14 criteria in the transcribed interviews. True cases included considerably more CBCA criteria (M = 7.63) than false cases (M = 4.08), and this difference was substantial, d = 2.67.

There are reasons to believe that applying the Validity Checklist is sometimes problematic. It is possible to question the justification of some of the external factors listed on the Validity Checklist, for example inappropriateness of affect. This refers to whether the affect displayed by the child when being interviewed (usually nonverbal behaviour) is inappropriate for the child’s alleged experiences. This external factor implies that the notion of appropriate affect displayed by victims of sexual abuse exists, whereas it does not. That is, in interviews, some sexually abused victims express distress that is clearly visible to outsiders, whereby others appear numbed and cues of distress are not clearly visible (Burgess, 1985; Vrij & Fischer, 1997). The communication styles represent a personality factor and are not related to deceit (Littmann & Szewczyk, 1983). Yet, emotional victims are more readily believed than victims who report their experience in a controlled manner (Baldry & Winkel, 1998; Baldry, Winkel, & Enthoven, 1997; Bollingmo, Wessel, Eilertsen, & Magnussen, 2008; Bothwell & Jalil, 1992; Hackett, Day, & Mohr, 2008; Kaufmann, Drevland, Wessel, Overskeid, & Magnussen, 2003; Rose, Nadler, & Clark, 2006; Vrij & Fischer, 1997; Wessel, Drevland, Eilertsen, & Magnussen, 2006).

Some other external factors are difficult to measure. Take for example susceptibility to suggestion. Statements of suggestible children could be problematic to interpret because suggestible children may be inclined to provide information that confirms the interviewer’s expectations but is, in fact, inaccurate. To examine a child’s susceptibility to suggestion,
the interviewer is recommended to ask the witness a few leading questions at the end of the interview (Landry & Brigham, 1992; Yuille, 1988). Interviewers should hereby only ask questions about irrelevant peripheral information, because asking questions about central information could damage the quality of the statement. Being allowed only to ask questions about peripheral information is problematic, as it may say little about the witness’ suggestibility regarding core issues of his or her statement. Children show more resistance to suggestibility for central parts than peripheral parts of an event (Dalton & Daneman, 2006; Davies, 1991; Goodman, Rudy, Bottoms, & Aman, 1990).

It is difficult, if not impossible, to determine the exact impact that many external factors have on CBCA scores. A good illustration is the field study conducted by Lamers-Winkelman and Buffing (1996). In this study, raters were instructed to take the age of the child into account when calculating CBCA scores. Nevertheless, six criteria positively correlated with age. In other words, even after being instructed to correct CBCA scores for age, the results still showed age-related effects with older children obtaining higher CBCA scores than younger children.

Given these difficulties in measuring the external factors and in examining the exact impact of these external factors on CBCA scores, it is clear that the Validity Checklist procedure is more subjective and less formalized than the CBCA procedure. It is therefore not surprising that if two experts disagree about the truthfulness of a statement in a German criminal case, they are likely to disagree about the likely impact of Validity Checklist external factors on that statement (Vrij, 2008, personal communication). One study revealed that Swedish experts sometimes use the Validity Checklist incorrectly, and this could be due to the difficulties with applying it (Gumpert & Lindblad, 1999). First, although SVA experts sometimes highlight the influence of Validity Checklist issues on children’s statements in general, they do not always discuss how these issues might influence the statement of the particular child they are asked to assess. Second, although experts sometimes indicate possible external influence on statements, they are inclined to rely upon the CBCA outcome, and tend to judge high-quality statements as truthful and low-quality statements as fabricated.

The latter, ignoring the Validity Checklist and solely relying on CBCA scores, is unfortunate. Most aspects mentioned on the Validity Checklist are relevant when making veracity judgements based on CBCA scores. For example, research has convincingly demonstrated that total CBCA scores are age dependent. With increased age, children obtain higher total CBCA scores (see Vrij (2008) for a list of 14 studies examining the CBCA–age relationship). The difficulties and problems
in using the Validity Checklist do not mean that these factors do not have an influence on children’s reports or that they should be dismissed.

**SVA Summary**

Although SVA assessments are used as evidence in (criminal) courts to evaluate the veracity of child witnesses’ testimonies in trials for sexual offences, the accuracy of these assessments is unknown. However, research has shown that CBCA-trained evaluators achieve around 70% accuracy. The Validity Checklist is difficult to apply for a variety of reasons, yet it is a vital element of SVA and should not be dismissed.

**REALITY MONITORING**

People sometimes try to determine whether they have actually experienced an event they have in mind, or whether this memory is based on imagination. The processes by which a person attributes a memory to an actual experience (external source) or imagination (internal source) is called Reality Monitoring (Johnson, Foley, Suengas, & Raye, 1988; Johnson & Raye, 1981). Although the RM concept is not related to deception, scholars believe that RM can be used as a lie detection tool and have examined this. To our knowledge, the RM lie detection tool is not used by practitioners.

The core of RM is that memories based on real experiences differ in quality from memories based on fiction. In their seminal work on memory characteristics, Marcia Johnson and Carol Raye (1981) argued that memories of real experiences are obtained through perceptual processes. They are therefore likely to contain *sensory information*: details of smell, taste or touch, visual details and details of sound; *contextual information*: spatial details (details about where the event took place and about how objects and people were situated in relation to each other) and temporal details (details about the time order and duration of events) and *affective information*: details about people's feelings throughout the event. These memories are usually clear, sharp and vivid. By contrast, memories about imagined events are derived from an internal source and are therefore likely to contain *cognitive operations*, such as thoughts and reasonings ('I must have had my coat on, as it was very cold that night'). They are usually vaguer and less concrete. There is empirical support for this general RM assumption. For example, in some studies, participants were asked to think about an
experienced event or about an imagined event. They were then asked to complete the Memory Characteristics Questionnaire which examines the quality of their memory. Participants in the real memory conditions had clearer memories that included more sensory information than participants in the imagined memory conditions (Gordon, Gerrig, & Franklin, 2009; Johnson et al., 1988).

Research has shown that people use these textual properties also to judge whether somebody else’s memory (Interpersonal RM) is externally or internally derived (Johnson, 2006; Johnson, Bush, & Mitchell, 1998; Johnson & Suengas, 1989). Indeed, one of the verbal criteria observers report they rely most upon when detecting lies is ‘richness in details’ (Strömwall, Granhag, & Hartwig, 2004; Vrij, 2008; Vrij, Akehurst, & Knight, 2006). The richer an account is perceived to be in detail, the more likely it is to be believed (Bell & Loftus, 1989; Johnson et al., 1988).

From 1990 onwards, scholars have examined whether RM analyses can be used to discriminate between truths and lies. The assumption those scholars make is that truths are recollections of experienced events whereas lies are recollections of imagined events. Obviously not all lies are descriptions of events that a person did not experience. Many lies are not about events, but are about people’s feelings, opinions or attitudes. And even when people lie about events (about their actions and whereabouts), they can sometimes describe events that they have actually experienced. For example, a burglar who denies having committed a burglary last night can claim that he went to the gym instead. He can then describe a truthful visit he made to the gym (but on another occasion).

Researchers have examined whether truthful statements about experienced events differ in terms of RM criteria from deceptive statements that are based on events that the liar imagined. The typical procedure is that liars and truth tellers are interviewed, and these interviews are taped and transcribed. RM experts check for the presence of RM criteria in these transcripts. A standardized set of RM deception criteria has not been developed to date, and different researchers use different criteria and sometimes use different definitions for the same criterion. However, most researchers include the following criteria in their RM veracity assessment tool (Vrij, 2008): **clarity and vividness** of the statement: is the report clear, sharp and vivid instead of dim and vague? **perceptual information**: the presence of sensory information in a statement, such as sounds (‘He really shouted at me’), smells (‘It smelled of rotten fish’), tastes (‘The chips were very salty’), physical sensations (‘It really hurt’) and visual details (‘I saw the nurse entering the ward’); **spatial information**: information about locations (‘It was in a park’) or the spatial arrangement of people and/
or objects (‘The man was sitting to the left of his wife’) and *temporal information*: information about when the event happened (‘It was early in the morning’) or explicitly described sequences of events (‘When he heard all that noise, the visitor became nervous and left’).

Other criteria that are part of the RM tool are *affect*: how the witness felt during the event (‘I was scared’), *reconstructability of the story*: is it possible to reconstruct the story based on the information given and *realism*: is the story plausible, realistic and does it make sense. The final RM criterion is *cognitive operations*: descriptions of inferences made by the participant at the time of the event (‘It appeared to me that she didn’t know the layout of the building’) or inferences/opinions made when describing the event (‘She looked smart’). All criteria are thought to be more present in truthful than in deceptive accounts, except for the cognitive operations criterion, which is thought to be present more in deceptive than in truthful accounts. The criteria are also presented in Table 1.2.

Reading the content of several CBCA and RM criteria suggests that there is some overlap between the CBCA list and RM criteria, a view empirically supported by Sporer (2004) who carried out correlational and factorial analyses on several data sets. The RM criteria *spatial*

<table>
<thead>
<tr>
<th>Table 1.2</th>
<th>The reality monitoring criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>Clarity and vividness of the statement</td>
</tr>
<tr>
<td>Sensory (perceptual) information</td>
<td>Details about what the interviewee saw, heard, smelled, touched or tasted.</td>
</tr>
<tr>
<td>Spatial detail</td>
<td>Information about locations or the spatial arrangement of people and/or objects.</td>
</tr>
<tr>
<td>Temporal detail</td>
<td>Information about when the event happened or explicitly describes a sequence of events.</td>
</tr>
<tr>
<td>Affect</td>
<td>Information about how the interviewee felt during the event.</td>
</tr>
<tr>
<td>Reconstructability of the story</td>
<td>Is it possible to reconstruct the event on the basis of the information that is given?</td>
</tr>
<tr>
<td>Realism</td>
<td>Is the story plausible, realistic and does it make sense?</td>
</tr>
<tr>
<td>Cognitive operations</td>
<td>Descriptions of inferences made by the interviewee at the time of the event or inferences/opinions made when describing the event. The presence of a cognitive operation suggests deceit.</td>
</tr>
</tbody>
</table>

Adapted with permission from Vrij (2008). © John Wiley & Sons, Ltd.
information and temporal information are related to CBCA Criterion 4, contextual embeddings. The difference is that spatial information and temporal information are two separate criteria in RM whereas they are combined into one criterion in CBCA. The RM criterion affect is related to CBCA Criterion 12, subjective mental state, although the CBCA criterion not only includes affect but also thoughts. The RM criterion realism is related to CBCA Criterion 1, logical structure. A difference, however, is that RM criterion realism takes plausibility into account whereas CBCA criterion logical structure does not. The sound detail in RM (part of the perceptual information criterion) is different from CBCA Criterion 6, reproduction of conversation. CBCA Criterion 6 is more restricted and only includes speech that is reported in its original form, whereas the RM criterion sound refers to sound in general. Thus the two phrases ‘He said: Are you OK, you look so pale?’ and ‘He asked how I felt’ both count as a sound detail in RM, whereas the latter phrase does not count as a conversation detail in CBCA because it is not verbatim. In addition, a car horn would count as a sound detail in RM but not as a conversation detail in CBCA.

**RM: Research**

Masip et al. (2005) and Vrij (2008) reviewed the RM deception research (all laboratory studies). Vrij (2008) reviewed the results of 30 samples and found that, in terms of individual criteria, in particular the idea that truth tellers recall more perceptual information and more spatial information and temporal information received support.

In 10 studies, RM was used to discriminate truth tellers from liars. The average truth accuracy rate was 72% and the average lie accuracy rate was 66% (Vrij, 2008), whereas 50% accuracy rate can be expected by chance alone. At least two more studies have been published since Vrij’s (2008) review which reported accuracy rates (Nahari et al., 2012, in press). These two studies obtained similar accuracy rates to the first 10 studies (71 and 63% total accuracy).

These RM truth and lie accuracy rates are similar to those obtained with CBCA. In fact, in most of those studies (N=8), researchers carried out RM and CBCA analyses, which makes a comparison of the two tools possible. The findings are inconclusive. In three studies, CBCA analyses resulted in superior total accuracy rates, and in the other five studies, the best accuracy rates were achieved with RM analyses. The average total accuracy rate for RM in those eight studies is slightly higher (68.13%) than the average total accuracy rate for CBCA (63.63%).

There are restrictions in using an RM veracity assessment tool. For example, the tool cannot be used with young children. In some
circumstances, children do not differentiate between fact and fantasy as clearly as adults do, for several reasons including that children have a richer imagination than adults (Lindsay, 2002). Children may therefore be better than adults at imagining themselves performing acts. It is probably also difficult to use the RM tool when people talk about events that happened a long time ago. Over time, cognitive operations may occur in memories of experienced events because they facilitate the remembering of events (Roediger, 1996). Someone who drove fast in a foreign country may try to remember this by remembering the actual speed his speedometer indicated; alternatively, he could remember this by logical reasoning and by deducing that he must have driven fast because he drove on the motorway. Imagined memories, on the other hand, can become more vivid and concrete over time if people try to visualize what might have happened (Manzanero & Diges, 1996).

**RM Summary**

RM is based on solid memory theory. It is popular amongst researchers, but not used in real life by practitioners. The RM tool achieves around 70% accuracy, which is similar to CBCA. RM cannot be used under all circumstances.

**SCIENTIFIC CONTENT ANALYSIS**

SCAN, developed by Avinoam Sapir, a former polygraph examiner in the Israeli police, is used worldwide and in countries such as Australia, Belgium, Canada, Israel, Mexico, the Netherlands, Qatar, Singapore, South Africa, the United Kingdom, and the United States (Vrij, 2008). It is used by federal law enforcement (including the Federal Bureau of Investigation (FBI)), military agencies (including the U.S. Army Military Intelligence), secret services (including the Central Intelligence Agency (CIA)) and other types of investigators (including social workers, lawyers, fire investigators and the American Society for Industrial Security) (Bockstaele, 2008; www.lsiscan.co.il). http://www.lsiscan.com/id29.htm provides a full list of past participants of SCAN courses. According to (the American version of) the SCAN website (www.lsiscan.com), SCAN courses are mostly given in the United States and Canada (on a weekly basis in those countries). In addition, online courses are also available.

In the SCAN procedure, the interviewee is requested to write a detailed description of all his/her activities during a critical period of time in such a way that a reader without background information can
determine what actually happened. The handwritten statement is then analysed by a SCAN expert on the basis of a list of criteria. It is thought that some SCAN criteria are more likely to occur in truthful statements than in deceptive statements, whereas other criteria are more likely to occur in deceptive statements than in truthful statements (Sapir, 1987/2000). However, despite its name, Scientific Content Analysis, SCAN is an a-theoretical and no theoretical justification is given as to why truth tellers and liars would differ from each other in the stated ways.

There is not a fixed list of SCAN criteria, and different experts seem to use different sets of criteria. The list of 12 criteria presented in Table 1.3 are those that are most emphasized in workshops on the technique (Driscoll, 1994), used in research (Smith, 2001) or were used frequently by SCAN users in a field observation (Bogaard, Meijer, Vrij, Broers, & Merckelbach, in press).

Denial of allegations refers to whether the interviewee directly denies the allegation in the statement. Denials are perceived as truthful. Social introduction refers to how the persons described in the statement are introduced. Honest social introductions are thought to be unambiguous (e.g. ‘My wife Lisa…’), whereas a failure to introduce someone (e.g. ‘We went outside’ without mentioning who ‘we’ are) is interpreted as the writer trying to hide something. It may also indicate tension between the people to whom the ambiguity refers. Spontaneous corrections refers to the presence of corrections in the statement, such as crossing out what has been written. Although explanations and additions are allowed, interviewees are explicitly instructed not to cross anything out. A failure to follow this instruction is believed to indicate deceit. Lack of conviction or memory is the interviewee being vague about certain elements in the statement (‘I believe…’, ‘I think…’, ‘kind of…’) or when the interviewee writes that he or she cannot remember something. SCAN users interpret these phrases as suspicious. Structure of the statement refers to the balance of the statement. It is thought that in a truthful statement the first 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. Thus, a 10-line statement is thought to comprise 2 lines to introduce the event, 5 lines to describe the event and 3 lines about the aftermath. The more unbalanced a statement, it is suggested the greater the probability that the statement is deceptive.

Emotions refers to whether there are emotions described in the statement. This criterion also refers to where the emotions are mentioned in the statement. It is thought that deceivers will mention emotions just before the climax of the story, whereas truth tellers are more likely to mention emotions throughout the story, but particularly after
Table 1.3  The scientific content analysis criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Denial of allegations</td>
<td>Does the interviewee directly deny the allegation in the statement? Truthful interviewees are more likely to do this.</td>
</tr>
<tr>
<td>2. Social introduction</td>
<td>How are the persons introduced who are described in the statement. They can be unambiguous (e.g. ‘My friend, by the name of Phil…’) or ambiguous (e.g. ‘We went outside’) without saying who ‘we’ are. Ambiguous introductions indicate a lie.</td>
</tr>
<tr>
<td>3. Spontaneous corrections</td>
<td>Corrections in statements, such as crossing out what has been written. Corrections are seen as sign of deceit.</td>
</tr>
<tr>
<td>4. Lack of conviction and memory</td>
<td>The interviewee is vague about certain elements in the statement (‘I think it was on the back of the chair but I just can’t remember’), or reports that she/he cannot remember something. Lack of memory indicates deceit.</td>
</tr>
<tr>
<td>5. Structure of statement</td>
<td>The balance of the statement. It is thought that in a truthful statement the first 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. The more unbalanced a statement, the greater the probability that the statement is deceptive.</td>
</tr>
<tr>
<td>6. Emotions</td>
<td>Description of emotions. Important is the position of the emotions in the statement: just before the climax of the story (liars), or throughout the story (truth tellers) and after the climax of the story (truth tellers).</td>
</tr>
<tr>
<td>7. Objective and subjective time</td>
<td>How different time periods are covered in a statement. Objective time is the actual duration of events described in the statement, whereas subjective time is the amount of words spent to describe these events. It is thought that in a truthful statement, the objective and subjective time will correspond with each other, unlike in a deceptive statement.</td>
</tr>
</tbody>
</table>

(Continued)
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Out-of-sequence and extraneous information</td>
<td>Does the statement recount the events in chronological order (out-of-sequence information)? A deviation of the chronological order indicates deceit. Also, does the statement contain information that does not seem relevant (extraneous information)? The presence may indicate deceit because interviewees could include extraneous information to hide more important information.</td>
<td></td>
</tr>
<tr>
<td>9. Missing information</td>
<td>The inclusion of words that indicate that some information has been left out, such as ‘sometime after’, ‘finally’, ‘later on’ and ‘shortly thereafter’. Inclusion of those words indicates deceit.</td>
<td></td>
</tr>
<tr>
<td>10. First person singular past tense</td>
<td>The format in which a statement is written. It is thought that truthful statements are written in the first person singular, past tense because the interviewee describes an event that has taken place (e.g. ‘At the time I just saw it happen, I just jumped down and picked the laptop up straight away’). Deviations from this norm (e.g. ‘It is the blue one’) when referring to a lost item should raise suspicion.</td>
<td></td>
</tr>
<tr>
<td>12. Change in language</td>
<td>Change of terminology or vocabulary in the statement. A change in language indicates that something has altered in the mind of the writer. For example, if an interviewee refers in his statement to all conversations he had as ‘conversations’ except one conversation which he describes as a ‘discussion’, it is likely that she/he perceived this conversation differently from the other conversations.</td>
<td></td>
</tr>
</tbody>
</table>

Adapted with permission from Vrij (2008). © John Wiley & Sons, Ltd.
the climax of the story. **Objective and subjective time** refers to how different time periods are covered in a statement. Objective time is the actual duration of events described in the statement, whereas subjective time is the amount of words spent describing these events. It is thought that in a truthful, but not in a deceptive statement, the objective and subjective time will correspond with each other. For example, if someone devotes five lines in a statement to describe a 30-minute period and then three lines to describe a subsequent 2-hour period, the objective and subjective time do not correspond and this may indicate deceit. **Out-of-sequence and extraneous information** refers first of all to whether the statement recounts the events in chronological order. A deviation of the chronological order may be deceptive. It also refers to extraneous information that does not seem relevant. It is thought that examinees could include extraneous information to hide more important information.

**Missing information** refers to phrases in the statement that indicate that some information has been left out. Examples are the use of words such as ‘sometime after’, ‘finally’, ‘later on’ and ‘shortly thereafter’. Missing information is seen as a sign of deceit. **First person, singular tense** refers to the format in which a statement is written. It is thought that truthful statements are written in the first person singular, past tense because the writer describes an event that has taken place (‘I saw the smoke coming out of the window’). Deviations from this norm are viewed with suspicion. **Pronouns** include words such as ‘I’, ‘my’, ‘he’, ‘his’, ‘they’ and ‘their’. Pronouns signal commitment, responsibility and possession. Omitting pronouns (‘Left the house’ rather than ‘I left the house’) suggests reluctance on the writer’s part to commit himself/herself to the described action. The use of ‘we’ when ‘I’ is appropriate is believed to suggest that the writer is trying to absolve himself/herself of personal responsibility. Leaving out pronouns that indicate possession (‘my’, etc.) suggests that the writer denies ownership. **Change of 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. For example, if a suspect describes in his statement all conversations he had as ‘conversations’ but one conversation as a ‘discussion’, it is considered likely that he perceived this conversation differently from the other conversations. When a change of language is noticed in a statement, a SCAN user should contemplate whether the sequence of events justifies such a change. If the SCAN expert can think of a justification, then she/he may consider that the interviewee may be truthful; if the expert cannot think of a justification, she/he should consider that the interviewee may be deceptive.
There is overlap between some SCAN criteria and some CBCA criteria although, intriguingly, SCAN and CBCA users often draw different conclusions about the veracity of a statement when a criterion is present. *Spontaneous corrections* is similar to CBCA Criterion 14, *spontaneous corrections*, but CBCA experts believe that spontaneous corrections indicate truthfulness whereas SCAN experts interpret it a sign to deceit. *Lack of conviction or memory* is similar to CBCA Criterion 15, *lack of memory*. Again, CBCA experts interpret lack of memory as a sign of truthfulness and SCAN experts as a sign of deceit. *Emotions* is similar to CBCA Criterion 12, *accounts of subjective mental state*. However, unlike CBCA, in SCAN, this criterion also refers to when the emotions are mentioned in the statement. *Out-of-sequence and extraneous information* is a combination of two CBCA criteria: *unstructured production* (Criterion 2) and *superfluous details* (Criterion 9). CBCA experts rate these criteria as signs of truthfulness and SCAN experts rate them as signs of deceit.

**SCAN: Research**

Research on SCAN is scarce. Given how often SCAN is used in real life, this is worrying and SCAN research is therefore urgent (Heydon, 2011). Only five SCAN studies have been published to date. In a field study, 30 written statements voluntarily given by suspects immediately prior to their polygraph tests were analysed (Driscoll, 1994). The SCAN expert allocated 73% of the truthful and 95% of the deceptive statements correctly. These accuracy scores are high and often mentioned by SCAN users. However, the study had an important limitation as acknowledged by the author of the paper himself. For each suspect, it is unknown whether he or she was actually telling the truth or lying. In technical terms, the *ground truth* was uncertain, which makes interpreting the results of this field study problematic.

A second SCAN field study was carried out in 2001 for the British Home Office (Smith, 2001). Three groups of SCAN users with different levels of experience, a fourth group of experienced detectives not trained in SCAN and a fifth group of newly recruited officers also not trained in SCAN analysed 27 real-life statements. As was the case in the first field study, ground truth was poorly established, which makes interpretation of the accuracy rates problematic. The three groups of SCAN users performed well and correctly classified at least 80% of the truths and 75% of the lies. However, so did the experienced detectives untrained in SCAN, and so the study therefore showed no benefit in using SCAN. The newly recruited officers performed worse than the SCAN groups and the experienced detectives but this may have little
to do with lie detection skills. Smith’s sample included only a small number of statements believed to be true and, as a result, a sceptical observer who had a lie bias and was inclined to judge a statement as deceptive was more likely to obtain a high accuracy rate. Research has indicated that training and experience tend to result in a lie bias (Meissner & Kassin, 2002), and the newly recruited officers were the least trained and experienced officers in the study.

Armistead (2011) challenged Smith’s (2001) data analyses and believes that Smith may not have come to the conclusion that SCAN users were as accurate as experienced lie detectors who are not trained in SCAN if she had analysed the data differently. He also criticized her decision to include infrequent SCAN users in the sample. According to Armistead, SCAN is not a template that can be implemented well without thorough practice. Armistead noted that a common experience of new graduates is to consult with Sapir himself or more experienced SCAN users for a substantial period of time after attending the SCAN training course. Smith’s field study also revealed that different SCAN experts used different SCAN criteria to justify their decision of whether or not a statement was deceptive. In other words, there was a lack of standardization in the application of SCAN amongst SCAN users. Lack of standardization is a serious problem, as it becomes unclear what the SCAN method actually entails: the way SCAN is used highly depends on the individual who applies it.

Porter and Yuille (1996) carried out a laboratory SCAN experiment in which participants did or did not commit a mock theft and were interviewed about this alleged crime. They examined three SCAN criteria: Structure of the statement, missing information and first person singular, past tense. Truthful and deceptive statements did not differ from each other on these three criteria. However, this experiment was also limited. First, it tested only a small number (three) of SCAN criteria. Second, participants gave oral statements that were subsequently transcribed. It means that the procedure did not follow one of the basic principles of SCAN, the request to write down a statement. In a fourth study, verbal criteria including some SCAN criteria (e.g. verb tense) were examined (Bachenko, Fitzpatrick, & Schonwetter, 2008). However, the results for the individual criteria were not discussed. In the final published SCAN study (Nahari et al., 2012), a laboratory experiment, truth tellers truthfully wrote down their activities during the last half hour, whereas liars reported a fabricated story. The statements were analysed with SCAN and, by way of comparison, also with RM. SCAN did not distinguish truth tellers from liars above the level of chance but RM did. With RM analyses, 71% of truth tellers and liars were correctly classified.
As mentioned above, there is some overlap between CBCA and SCAN criteria. For example, the criteria spontaneous corrections, lack of memory, out-of-sequence information and extraneous information appear on both lists, but the predictions about how these criteria discriminate between truth tellers and liars differ. In CBCA, the occurrence of those cues is perceived as an indicator of truth, whereas in SCAN, the same criterion is seen as an indicator of deceit. There is a substantial amount of CBCA research regarding these individual criteria (all laboratory studies) and they all give support only to the CBCA assumptions (Vrij, 2008).

To summarize SCAN research, two field studies with poor ground truth have been published. The findings of one study were favourable for SCAN, but the findings of the other were not, as the same high performance was achieved by non-SCAN users. That study also showed a lack of standardization in the way users apply SCAN. The SCAN laboratory experiments and the CBCA laboratory experiments that examined some individual SCAN criteria showed no support for SCAN.

SCAN Summary

SCAN is popular amongst practitioners and is widely used. There is not much research into it, but research carried out to date has shown that it lacks standardization and that it is subjective in that how it is used depends on the individual SCAN user. In addition, research has yet to demonstrate that SCAN actually works.

A SVA, RM AND SCAN COMPARISON

This section compares the three verbal veracity tools discussed in this chapter. To give structure to this comparison, the three tools are discussed by addressing the five criteria that are used by the United States Supreme Court for admitting expert evidence in American federal courts. The five criteria together address the theoretical underpinning and accuracy of the tools as well as the view of the scientific community about the tools. The following five criteria were used in the Daubert case (Honts, 1994): (1) Is the scientific hypothesis testable? (2) Has the proposition been tested? (3) Is there a known error rate? (4) Has the hypothesis and/or technique been subjected to peer review and publication? and (5) Is the theory upon which the hypothesis and/or technique is based generally accepted in the appropriate scientific community? In this section, one question was added to this list which
Verbal Lie Detection Tools

Table 1.4  A summary of the three tools based on the Daubert guidelines

<table>
<thead>
<tr>
<th></th>
<th>CBCA</th>
<th>RM</th>
<th>SCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Is the scientific hypothesis testable?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(2) Has the proposition been tested?</td>
<td>Yes</td>
<td>Yes</td>
<td>Scarcely</td>
</tr>
<tr>
<td>(3) Is there a known error rate? a</td>
<td>Yes, 30%</td>
<td>Yes, 30%</td>
<td>No</td>
</tr>
<tr>
<td>(4) Has the hypothesis and/or technique been subjected to peer review and publication?</td>
<td>Yes</td>
<td>Yes</td>
<td>Barely</td>
</tr>
<tr>
<td>(5) Has research supported the hypothesis and/or technique?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>(6) Is the theory upon which the hypothesis and/or technique is based generally accepted in the appropriate scientific community?</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Adapted with permission from Vrij (2008). © John Wiley & Sons, Ltd.
aBased on laboratory studies only.

seems important in a section that compares the three tools: has research supported the hypothesis and/or technique? Table 1.4 summarizes the answers to these questions for SVA, RM and SCAN assessments.

**Question 1: Is the Scientific Hypothesis Testable?**

The prediction that truthful statements will differ in quality from false statements because liars have difficulty in fabricating information and attempt to avoid looking suspicious (the rationale underlying CBCA) can easily be tested in laboratory research. The same applies to the underlying assumption of RM that memories based on real experiences differ in quality from memories based on fiction, which results in verbal differences between truth tellers and liars. SCAN has no theoretical rationale, and no justification is given as to why the SCAN criteria would differentiate truth tellers from liars. However, the hypothesis that truth tellers and liars do differ from each other in terms of SCAN criteria can be tested. Although testing the underlying rationales in field studies is possible in principle, in reality, it is difficult given the problems with establishing the truth or falsity of statements beyond doubt (ground truth). The answers are therefore ‘yes’ for CBCA, RM and SCAN.

**Question 2: Has the Proposition Been Tested?**

The answer to the second *Daubert* question is affirmative for CBCA and RM research. A substantial number of CBCA and RM studies have
been carried out, although in most studies adults rather than children participated (this is relevant for CBCA as that tool is designed for use with children). There have only been a few CBCA field studies conducted and several of those are of poor quality. No field studies exist that focus on the entire set of RM criteria. There are only five SCAN studies published to date and most of those have serious limitations. Although several CBCA studies have examined criteria that are also on the SCAN list, the SCAN tool as a whole has been scarcely examined to date. The answer for SCAN as a whole is thus ‘scarcely’.

Question 3: Is There a Known Error Rate?  
There are known error rates for CBCA and RM judgements made in laboratory research. These error rates are about 30% for both tools. An error rate for SCAN is not available. As mentioned earlier, SVA assessments are used as evidence in criminal courts in several European countries. The standard of proof set in those courts is ‘beyond reasonable doubt’. An error rate of about 30% in that context is too high.

Question 4: Has the Hypothesis and/or Technique Been Subjected to Peer Review and Publication?  
A growing number of CBCA and RM studies have been published in peer-reviewed journals, both laboratory studies and (for CBCA) field studies. Although some CBCA studies have examined some SCAN criteria, peer-reviewed studies about SCAN as a whole are almost lacking. The answer to the fourth Daubert question is thus ‘yes’ for CBCA and RM research, and ‘barely’ for SCAN research.

Question 5: Has Research Supported the Hypothesis and/or Technique?  
Empirical studies provide general support for CBCA and RM, and the answer for CBCA and RM is therefore ‘yes’. SCAN studies are scarce and, if we disregard the field study that lacked ground truth, do not support SCAN. The CBCA studies that examined SCAN criteria do not support the SCAN assumptions either. The answer is therefore ‘no’ for SCAN.

Question 6: Is the Theory upon Which the Hypothesis and/or Technique is Based Generally Accepted in the Appropriate Scientific Community?  
Several authors have expressed serious doubts about the SVA method (Brigham, 1999; Davies, 2001; Lamb et al., 1997; Pezdek & Taylor, 2000; Rassin, 1999; Ruby & Brigham, 1997; Wells & Loftus, 1991). For example, Lamb et al. (1997, p. 262) who carried out a reliable field study
with the ground truth being well established concluded that ‘... the level of precision clearly remains too poor to permit the designation of CBCA as a reliable and valid test suitable for the courtroom’. However, since then two CBCA field studies have been published which showed that CBCA analyses are a valuable tool in assessing child sexual abuse cases. Authors have also been critical about SCAN; see, for example, Heydon (2008, 2011), Nahari et al. (2012), Smith (2001), Verschuere (2008a, 2008b) and Verschuere, Meijer, and Vrij (2010). Scholars’ views on RM seem to be more positive, and critical articles about the use of RM as a lie detection tool have not yet emerged. Of course, RM is not used by practitioners in the field and it may be that scholars who are sceptical about the technique do not feel the urge to criticize it. To answer this Daubert question properly, one needs to know the opinion of the appropriate scientific community. To gather the views of the scientific community, surveys amongst scholars are required, as has been done about physiological (polygraph) lie detection (Iacono & Lykken, 1997). The appropriate scientific community has not been consulted about verbal lie detection tools to date. The answer to this Daubert question is thus ‘unknown’ for all tools.

Final Verdict

The known CBCA error rate (based on laboratory studies only) is 30%, which is not beyond reasonable doubt, the standard of proof regularly set in criminal courts. This error rate, which amounts to almost a one in three chance of being wrong, makes me conclude that SVA assessments should not be allowed as evidence in criminal courts. I am aware that this firm standpoint will be challenged by SVA proponents. Their main argument is that SVA assessments need to be compared with other evidence presented in criminal courts that may be even less accurate (Köhnken, 2004). Indeed, if judges and jurors make intuitive decisions, the error rate is in all likelihood higher. A meta-analysis of lie detection performance by naive observers showed that they obtained on average a 54% accuracy rate, which represents a 46% error rate, considerably higher than the CBCA error rate. Therefore, replacing CBCA assessments with intuitive judgements would make matters considerably worse.

I expect that more people working in the criminal justice system will be sensitive to the ‘If not SVA, what is the alternative?’ argument. However, there are arguments against this point of view. For example, in all likelihood, judges, jurors and defendants will perceive the evidence that SVA experts present in court as strong. In cases where the defendant is innocent but the SVA expert considers the accusation
made against the defendant to be truthful, this may easily lead to a guilty jury verdict and, as a result, a wrongful conviction. It could also lead to a false confession if a defendant comes to the conclusion that the SVA expert’s verdict means that she/he can no longer avoid conviction. Defendants sometimes falsely confess if they believe that a conviction is certain because a confession may lead to a more lenient sentence (Vrij, 2008). Therefore, in my view, if the evidence presented in a criminal court is likely to be influential, a high accuracy level is necessary. If SVA assessments are accepted as evidence in criminal courts, then, at the very least, SVA experts should present the problems and limitations associated with their assessments when giving their expert testimony, so that judges, jurors, prosecutors and solicitors can make a considered decision about the validity of their expert testimony. These considerations do not apply to RM and SCAN because RM/SCAN assessments are not introduced as evidence in court.

The answers to the six questions do not really differ from each other for CBCA and RM and the two tools are comparable. A possible advantage of RM is that it is based on sound theory and that it has a stronger theoretical rationale than CBCA. Another advantage of RM is that it is quicker to use and easier to teach (Vrij, 2008). However, in terms of accuracy, CBCA and RM are probably similar. SCAN is the most problematic tool of the three. It is not based on any theory which is in itself problematic. A theory can explain why a tool works and when it works. Without an underlying theory, such questions cannot be answered. SCAN research has hardly been carried out, and the study SCAN users cite as support (Driscoll, 1994) is of poor quality. Other, more experimentally sound studies found no support for the SCAN tool (Nahari et al., 2012) or contradict the SCAN assumptions (many CBCA studies). Given these findings, it is worrying that SCAN is endorsed and used by so many practitioners.

REFERENCES


Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.


