

An Evaluation of an Abbreviated Version of a Checklist to Aid Judgements of Credibility in the Medico-Legal Setting.

Simon Easton, Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth, PO1 2DY. UK.
Email: simon.easton@port.ac.uk

Lucy Akehurst, Department of Psychology, University of Portsmouth, UK.
Liam Satchell, Department of Psychology, University of Portsmouth, UK.
Sarah Turtle, Department of Psychology, University of Portsmouth, UK.

This study investigated a 9-item version of an existing 28-item checklist which improved interviewers' classification of interviewees who were truth tellers and interviewees who were exaggerating their symptoms. Eighty-eight adult participants made a credibility judgement based on a video-recorded interview of subjects following an ice plunge experience. Evaluators, using an abbreviated checklist or no checklist, performed at a level no better than chance. Those using the full checklist achieved an overall hit rate significantly higher than chance level (86% overall accuracy with 100% accuracy for truth tellers and 73% accuracy for exaggerators). Results confirmed that the original checklist improved the ability of evaluators to distinguish between truth tellers and exaggerators, while reduction of the checklist reduced its utility. Lens modelling highlighted the five items that were most able to aid in the accurate detection of a malingering interviewee.

Keywords: deception, malingering, medico-legal reports, expert witnesses

Expert witnesses in court need to convince the judge that their opinions are well founded, drawing upon factual evidence, which requires expertise in its interpretation and presentation, as well as hearsay evidence such as presentation of the consensus of medical opinion on the causation of particular symptoms or conditions.

Expression of opinion as to the nature and cause of an individual's complaints requires that the expert identify evidence or assumptions upon which an opinion is based. When an expert witness provides an opinion based on the balance of probabilities, the primary focus is typically on the cause and course of the difficulty reported. However, there is a second focus: an implicit or explicit evaluation of the validity of the interviewee's reporting. Thus, for example, an expert may express a belief that an individual probably experienced a severe post-trauma reaction, but he/she will also need to consider whether the interviewee is exaggerating the consequences of the event.

Expert witnesses in the medico-legal setting should be highly trained in the assessment and detection of the existence or otherwise of complaints, be they physical or psychological. However, they are less likely to be highly trained in distinguishing between truth telling from exaggerating and malingering.

Easton, S., Akehurst, L., Satchell, L., & Turtle, S., (2017). An evaluation of an abbreviated version of a checklist to aid judgements of credibility in the medico-legal setting. *Open Access Journal of Forensic Psychology*, 9, 1-18.

The expert witnesses' knowledge of the literatures on lying, deception, and malingering is unlikely to match their clinical knowledge (Drob, Meehan, & Waxman, 2009; Hall & Hall, 2006; Kramer & Gagliardi, 2009; Resnick, 1995; Rogers, 1997; 2008). Yet, it could be argued that the absence of informed assessment of the validity of the interviewee's presentation at interview could seriously detract from the Court's ability to place confidence in other opinions expressed about that person's condition and reported experience.

It can be argued that all opinion should be provided with qualification, especially as there is much evidence to suggest that the detection of lying is not as easy as many would expect (Vrij, 2008; Vrij, Akehurst, Brown, & Mann, 2006; Vrij, Granhag, & Porter, 2010). Most legal professionals and skilled police/parole officers tend to perform at the same level as most lay people, i.e. at chance level (Bond & DePaulo, 2006; Porter, Woodworth, & Birt, 2000). However, being an expert witness or police officer may increase self-confidence in one's ability to detect deceit, without actually increasing the accuracy of that ability (Vrij, Akehurst, & Knight, 2006; Vrij, Akehurst, Soukara, & Bull, 2004a; Vrij, Edward, & Bull, 2001; Faust and Ahern, 2012).

A number of procedures have been developed in an endeavour to help experts distinguish between truthful and false accounts of events in the forensic setting. Statement Validity Assessment (SVA) provides a systematic credibility assessment of verbatim accounts relating to interviewees' reporting of events and experiences. Criteria-Based Content Analysis (CBCA), one part of SVA, seeks to direct the attention of the interviewer towards aspects which the authors believe are synonymous with truthful accounts (e.g. descriptions of interactions and reproduction of conversation).

There is evidence that assessors using CBCA differentiate between truthful and deceptive accounts under certain conditions with a degree of success which is better than chance (Blandón-Gitlin, Pezdek, Lindsay, & Hagen, 2009).

A second credibility assessment tool, Reality Monitoring (RM), examines the content and quality of information provided by interviewees (Johnson & Raye, 1981). RM proposes that recollections based on true experiences, more so than fabricated accounts, will include perceptual information such as visual details and physical feelings, information regarding when and where an event happened, and descriptions of emotional experience. RM suggests that fabricated accounts will involve more details of thoughts, reasoning, and inferences (e.g.; "I must have slept really poorly for weeks because my leg was so painful").

Whilst there is limited empirical evidence in support of RM, and, indeed, some contradictory findings, a review of the literature concluded that RM did tend to discriminate above chance level between truth tellers and liars (Masip, Sporer, Garrido, & Herrero, 2005). In a review of ten studies, RM showed an average accuracy rate for identifying truth tellers of 71.7%, and for liars of 66.1% (Vrij, 2008). Several studies have indicated that the combined accuracy of RM and CBCA criteria tend to be superior to that for CBCA criteria alone (Sporer, 1997; Strömwall, Bengtsson, Leander, & Granhag, 2004; Vrij, Akehurst, Soukara, & Bull, 2004b).

The forensic psychology literature on lying has tended to focus on a dichotomous conception of honesty and dishonesty. In medico-legal settings, however, it may be the case that partial truths will add to the complexity of the picture. Thus, it has been suggested that a distinction be made between feigning non-existent disease (Pure malingering); exaggeration of existing symptoms (Partial malingering); and the false ascribing of real symptoms to unrelated causes (False imputation) (Resnick, 1995). Faust, Ahern, Bridges & Yonce (2012) have proposed four possible categories for combinations of accurate and inaccurate reporting by assessors, wherein they identify the possibility that those being assessed may be neither malingering nor injured, and judgements about both aspects are correct, or incorrect, or judgements about either may be incorrect. They point out that the existence of the four possible responses leads to chance level of 25% and error rate of 75%, this illustrating that dual identification magnifies the likelihood of errors in assessment.

Malingering or exaggerating distress or difficulty, rather than simple lying, might affect both cause and course of difficulties. Various factors such as blame, expectations, labelling, focus on complaints, and social factors might affect an individual's experience and reporting of difficulties after an accident (Ferrari, Kwan, Russell, Pearce & Schrader, 1999). Patterns of symptoms have also been identified as warranting careful examination, with attention to reporting of improbable symptoms or combinations of symptoms in the context, or complaints of symptoms of improbably extreme severity (Rogers, 1997; Miller, 2001; Resnick & Knoll, 2005).

In their endeavours to improve the reliability of their assessments, medico-legal experts will try to conduct comprehensive assessments, and draw upon psychometric and other objective aids for assessment, where relevant. While psychological tests are promoted in some circumstances and settings as relevant and potentially valid (e.g. neuropsychological evaluations), interpretation of results can be problematic. For example, individuals inventing a history of pain may produce markedly similar scores on psychometric tests to those of pain patients, and people exaggerating existing pain tend to substantially over-endorse symptoms (Maguire, Harvey, & Shores, 2001). Further, studies have shown that successful malingerers/exaggerators, when asked to feign psychological difficulties, tend to endorse a lower rate of legitimate symptoms, avoid overly unusual or bizarre symptoms, and base their responses on their personal experiences. These are the kinds of difficulty facing assessors in the medico-legal setting that make the task of identifying false reporting difficult in many cases (Edens, Otto, Buffington, Tomicic, & Poythress, 2001).

In the USA, various structured assessments such as the Structured Interview of Reported Symptoms – 2nd Edition (SIRS-2; Rogers, Sewell, & Gillard, 2010) and brief screening measures such as the Miller Forensic Assessment of Symptoms Test (M-Fast; Miller, 2001, 2004) have been used, although concerns have been expressed about reliance on such devices in the medico-legal setting (Rogers et al., 2010). In the UK, such assessments are rarely used, however, perhaps because of differences in the legal process or higher levels of concern about the reliability and validity of such devices (Zapf & Grisso, 2012).

Those trying to detect dishonest reporting need to carefully consider an interviewee's non-verbal behaviour, examine physiological responses, and evaluate what a person says (Vrij, 2008). In practice, medico-legal experts will informally evaluate the interviewee's presentation to a greater or lesser degree. That informal evaluation will draw upon an expert witness's knowledge of the

relevant literatures into malingering and deception or lying. That implicit reference to the literature will be limited both by the familiarity of the expert with the relevant material and by the expert's ability to recall and consider that knowledge in the assessment setting.

Akehurst et al., 2015, drew upon the separate literatures into lying and malingering to develop an aide-memoire in the form of a checklist of items to direct interviewers' attention to valid cues to credibility and malingering. Following evaluation of the checklist, Akehurst et al., (2015), reported that evaluators without the checklist correctly identified 59% of truthful interviewees and 56% of interviewees who were exaggerating their symptoms following an ice plunge experience. Overall, results showed that the control group of evaluators classified interviewees at a level no better than chance (50%). However, the evaluators who used the checklist classified their interviewees at a level significantly better than chance (70%, $p < 0.01$), achieving a 75% correct classification of truth tellers and a 66% correct classification of interviewees exaggerating their symptoms. Further analysis of the data identified nine checklist items that most reliably discriminated between truth tellers and exaggerators, of which the majority were drawn from the work on malingering of Rogers (2008) and Ferrari et al., (1999).

While the checklist is brief, if such a tool is to be used in the time-pressured medico-legal setting, it will need to be shown to be practical, as well as providing statistically significant levels of discrimination, and will need to be as succinct and as easy to complete as possible.

The analysis by Akehurst et al. (2015) indicated that there might be an opportunity to reduce item redundancy, whilst maintaining the effectiveness of the checklist as a means of accurately judging credibility. The current study was therefore undertaken in the context of awareness of the time pressure on expert witnesses, noting that the questionnaire would need to be as efficient as possible if it is to be adopted in the applied setting. Thus, for example, Hinkin (1995) has noted that adequate internal consistency/reliability can be provided by as few as three items, and emphasised that, sometimes thereafter, additional items have progressively less impact on scale reliability.

This research addressed the possibility that a substantially shorter checklist than that which was tested previously (Akehurst et al., 2015), might maintain high levels of discrimination between interviewees telling the truth about their symptoms and those who were exaggerating their symptoms, and reduce the time involved in using such a device in the time-pressured medico-legal environment. In the current study, evaluators were given a 9-item checklist comprising the most successful items as highlighted by Akehurst et al. (2015), and their performance for accurately identifying truth tellers and exaggerators was compared to that of a group who received the original 28-item checklist and a group who received no checklist.

Method

Participants

Participants were recruited via advertisements to students and staff at a University in the UK. A small payment of £5 was made to each attendee. The 88 participants (44 females, $M_{Age}=27.4$ years, $SD_{Age}=9.6$ years, $Min_{Age}=19$ years, $Max_{Age}=58$ years) were evenly balanced across all conditions.

All participants were undergraduate students in the first term of their psychology degrees and none had any medical or legal training.

Materials (for more detail see Akehurst et al. 2015)

Cold pressor procedure. All interviewees¹ were asked to take part in a cold pressor procedure that would constitute the event that they would later be interviewed about. A cold pressor apparatus was used consisting of an insulated container filled with water and ice. Participants were instructed to keep their forearm and hand immersed for as long as they felt they could tolerate, with a maximum period of immersion of 3 minutes. They were then given directions to tell the truth or exaggerate their experience to those interviewing them.

Truth tellers. Those in the truth-telling condition were told to “Tell the truth about your experience. Try to include lots of truthful detail about your physical and psychological symptoms to make a really credible impression. You might imagine that, should the interviewer believe you, you would be awarded monetary compensation.” The participants were then given five minutes to prepare their accounts.

Exaggerators. Exaggerators were given the following instruction, “We would like you to convince the interviewer that the discomfort you experienced was worse than it actually was. You are also asked to pretend that the symptoms are ongoing and substantial. Try to include lots of detail about your physical and psychological symptoms to make a really credible impression. You might imagine that, should the interviewer believe you, you would be awarded monetary compensation.” These participants were then given a list of items to consider when preparing to exaggerate symptoms². These included suggestions to think about the type of pain in their hand, the location of the pain, the colour of their hand, the mobility of their hand, the possibility that they might have experienced psychological symptoms such as panic, anger, and anxiety and also the possibility of experiencing physiological symptoms such as dizziness and nausea. None of the suggested symptoms related directly to the items on the checklist. All interviewees were given five minutes to prepare their accounts.

The interviews. Truth tellers and exaggerators were interviewed by one of four interviewers and were all asked the same questions: “Please could you tell me everything you can about the experience of plunging your arm into very cold water? Can you tell me about the symptoms that you experienced whilst you had your arm in the cold water? Can you tell me about the symptoms you experienced once you had taken your arm out of the cold water? Are you still experiencing any symptoms? If yes, could you tell me about the symptoms that you are still experiencing? Now

¹ Of the interviewees, 32 (24 females, $M_{Age}=21.97$ years, $SD_{Age}=4.23$ years, $Min_{Age}=18$ years, $Max_{Age}=38$ years) were asked to tell the truth about their symptoms following the cold pressor procedure, and 32 (17 females, $M_{Age}=26.03$ years, $SD_{Age}=9.71$ years, $Min_{Age}=19$ years, $Max_{Age}=61$ years) were asked to exaggerate their symptoms.

² Truth tellers and exaggerators were treated differently to mimic real life. A list of considerations was provided to exaggerators to replicate the preparation that it is likely that those in real life would undertake before attending a meeting during which they intend to lie about their symptoms. Those malingering may resort to websites and other medical literature to help them to concoct a plausible account of non-existent or exaggerated symptoms. Truth tellers, on the other hand, tend to believe in the illusion of transparency i.e. that the truth will ‘shine through’ and so will simply be confident to report their true symptoms with little preparation pre-interview.

please tell me once again about your experience of putting your arm in the ice water and any subsequent discomfort up to the time of this interview. Please include as much detail as you can.” All interviews were video recorded. The interviewers did not appear in the shot; however the interviewees were clearly visible, with an unobstructed view from the top of their heads to their feet.

The checklist. Akehurst et al. (2015) developed the checklist that formed the basis of this experiment. Some 49 items drawn from a review of the relevant literature, which included 19 items from CBCA, 11 from the Validity Checklist of SVA, 8 RM criteria, 2 paralinguistic cues identified as useful (DePaulo, et al., 2003; Vrij, 2008), and 9 items from the malingering literature (Ferrari et al., 1999; Resnick, 1997; Rogers, 1997) were reduced to 28 items which were transferable to the medico-legal setting and could be used in an experimental setting.

There was a scale for each checklist item where, dependent on whether the literature suggested the item should appear more often in truthful or exaggerated accounts, sometimes 1 = not at all present and 5 = very much present, and for other items 1 = very much present and 5 = not at all present. Thus, the higher the score for each item, the more likely it was that the interviewee was exaggerating. For example, the literature suggests that, if an interviewee blames someone else or considers the other party was reckless, he/she is more likely to be lying/exaggerating. So, for this item, the scale ran from 1 = not at all present to 5 = very much present. Conversely, research suggests that the CBCA criteria are more likely to be present in truthful accounts. So, for these items, (e.g. unusual details, reproduction of conversation, quantity of detail) the scale ran from 5 = not at all present to 1 = very much present. The interviewers were asked to add up their item scores to give a total, and were instructed that the higher the total for the checklist, the more likely that their interviewee was exaggerating his/her symptoms. The full checklist can be seen in Appendix 1, which also identifies the source of the items and the nine items that were included in the abbreviated version.

The abbreviated checklist. To determine whether total checklist scores and scores for the individual items on the 28-item checklist were predictive of objective truth status, Akehurst et al. (2015) conducted a discriminant analysis with actual truth status of interviewees as the dependent grouping variable and the individual criteria scores as the predictor variables. The items that significantly discriminated between truth-tellers and exaggerators, and were therefore retained in the abbreviated version, were (i) details about the interviewee’s mental state, (ii) unusual details, (iii) attribution of blame, (iv) attribution of recklessness, (v) suggestion that symptoms only lasted a short period of time, (vi) confirmation of many symptoms, (vii) unexpected symptoms, (viii) contradictory symptoms and, (ix) extreme symptoms.

Design

A between-subjects design was used. One group of evaluators used an abbreviated checklist based on the nine checklist items which Akehurst et al (2015) found to significantly discriminate between truth tellers and exaggerators, one group used the full checklist (28 items) from the previous study, and one group was not provided with a checklist. The evaluators without the checklist were asked

to simply indicate whether their interviewee was truthful or exaggerating. All evaluators were additionally asked to rate their confidence in their judgement.

Evaluators watched one video recording of an interview. There were 64 videos in total with participant 1 being presented with video 1, participant 2 being presented with video 2, and so on. Participant 65 saw video 1 and the cycle repeated. Half of the evaluators from each of the three conditions watched a truth teller and half watched a malingerer. Evaluators were alternatively assigned to each condition, with participants arriving to the experiment in a random order. The evaluators were informed that it was equally likely that they would watch a truth teller as it was that they would watch a malingerer. Each recorded interview was approximately five minutes in length and was played only once

Procedure

All evaluators were given the following instructions: “Our interviewees have undergone a cold pressor procedure which involved them holding their forearm and hand in very cold water. We asked some people to exaggerate the discomfort that they experienced, while others were simply telling the truth. We want you to decide whether your interviewee is telling the truth or exaggerating.”

The evaluators who were given a checklist (either the full version or the abbreviated version) were also told: “Here also is a checklist that the research team has devised. This checklist is made up of items which are thought to be indicators of truthfulness or lying. Before you make your final credibility judgement, work through this checklist and keep your responses in mind when you make your final decision about your interviewee. Circle a number for each item and then add up to get a total score. The higher the total score the more likely that your interviewee is exaggerating his/her symptoms.”

None of the evaluators who were given the full or abbreviated checklist were given any further guidance regarding how to use them.

Results

Accuracy Rates

Evaluators who were given the full 28-item checklist achieved an overall hit rate of 86% (100% for truth-tellers and 73% for exaggerators), which was significantly higher than chance level (binomial test at 50%, $p=.001$). Those who were given an abbreviated version of the original checklist did not correctly classify their interviewees at a level significantly better than chance (overall accuracy 64%; accuracy for truth-tellers, 91%, and for exaggerators, 36%, binomial test at 50%, $p=.286$). Similarly those evaluators who were given no checklist did not perform at a level significantly above that expected by chance (overall accuracy 64%; accuracy for truth-tellers 73%, and for exaggerators 55%, binomial test at 50%, $p=.286$). These results demonstrated the effectiveness of the full checklist and highlighted that abbreviating the checklist hindered accuracy.

Confidence

There was no difference, across the three conditions, in the confidence evaluators had in their judgements ($F(2,63)=0.01, p=.991, f^2=.00$). It is notable that, on the 1 (not at all confident) to 7 (extremely confident) Likert scales, the evaluators generally lacked confidence in their judgements ($M_{Confidence}=2.76, SD_{Confidence}=1.28$). Evaluators who had one of the two different forms of the checklist did not differ in their ratings of the helpfulness of the checklist ($t(42)=0.39, p=.700, d=.12$) nor the degree to which the checklist agreed with their instinctive judgement ($t(42)=0.20, p=.842, d=.06$).

Item Utility

Further analysis was conducted to explore how well the individual items of the checklists indexed the veracity of the interviewees and how influential individual items were to the overall judgements made by evaluators. This was addressed by correlating the evaluators' coding of the checklist items with the veracity of the interviewees (truth teller or malingerer) and also with the judgement of the participant (truth or exaggeration). Ratings made by evaluators in both the full and abbreviated checklist conditions were included in the analyses to best describe the utility of each item. This form of analysis (adopted from lens modelling, see Karelaia & Hogarth, 2008; and for a deception detection context; Hartwig & Bond, 2011) highlighted the items most able to aid the accurate detection of a malingering interviewee (demonstrated by a negative correlation, see column 2 of Table 1). The analysis also identified the items most likely to influence a judgement of exaggeration (as demonstrated by a negative correlation, see column 3 of Table 1). The size of the correlations demonstrate the sensitivity of a particular item to aid the accurate detection of malingering and the influence that an item had on participants' judgements.

There were some discrepancies between those items which were strong indicators of veracity and those items which strongly influenced judgements. In other words, the items that were more influential to participants' judgements were not necessarily the most useful items. That is, the strongest correlations (all above the notable correlation value of .40 [Fergusson, 2009]) tended to either appear in the 'veracity' column for an item *or* in the 'judgement' column. Strong correlations rarely appeared in both the 'veracity' and 'judgement' columns for an item. Exceptions to this included the items: '*Does the individual believe that the other person behaved recklessly, so causing their discomfort?*'; '*Does the individual show long latency periods, i.e. the time between the end of a question and the beginning of an answer?*'; '*Does the individual show cognitive operations: inferences at the time of the event or at interview? (e.g. "It appeared to me that she was enjoying me doing the ice plunge")*', which are indicators of malingering, as well as; '*Does the individual expect his/her discomfort to last a short time and/or show significant improvement?*' and '*Does the individual provide temporal details? (e.g.; details about time order of the events; "the pins and needles came before the numb feeling")*' which are indicators of credibility. These five items were shown to accurately predict interviewee veracity and influenced evaluators' credibility judgements. Furthermore, we would like to highlight that the 'mismatch' between the accuracy and utilisation of the other 23 items were not significant (using Fisher's z tests to compare the correlations in the 'veracity' column and the 'judgement' column $M_z=.60, SD_z=.36, Min_z=.03, Max_z=1.68, all p=>.093$), which suggests there was not such a great 'mismatch' that participants incorrectly used the items to inform their judgements. The correlations of veracity indicators and the correlations of judgement indicators were all in the same direction; suggesting that, regardless of strength, an item which was more indicative of malingering was considered as such when

making overall judgements of credibility (the only exception was the item *Spontaneous Corrections*).

The results showed the benefit of the items on the checklist as a whole, but it seems that evaluators preferred to be informed by some items that were not the strongest indicators of accuracy.

Table 1. Correlations demonstrating how the evaluators' coding reflected the interviewees' veracity and how well the evaluators used their coding to inform their judgements.

Item	As an indication of interviewee veracity		As an indication of participant judgements	
	r	N	r	N
Blame	-.33	28	-.52	28
Reckless	-.44	22	-.58	22
Discomfort	-.57	37	-.45	37
Many Symptoms	-.31	43	-.40	43
Are not expected	-.11	43	-.21	43
Contradictory	-.21	43	-.31	44
Improbable	-.23	44	-.54	44
Unusual Details	.20	43	.38	44
Mental State	.13	43	.24	43
Falsely Ascribe	-.05	19	-.34	19
Lack Cooperation	.22	22	.17	22
Disorganised	.20	22	.34	22
Quantity of Detail	-.46	22	-.15	22
Spontaneous Corrections	-.12	21	.16	21
Raising Doubts	.33	20	.44	20
Physical Sensation	-.21	21	-.29	21
Temporal Details	.61	22	.55	22
Duration	.11	22	.30	22
Long Pause	-.39	21	-.40	21
Long Latency	-.40	21	-.44	21
Long Answer	.18	21	.43	21
Plausible	-.38	21	-.32	21
Superfluous Detail	.20	22	.44	22
Unexpected Complications	.24	21	.50	21
Avoids Lack of Memory	-.22	18	-.47	18
Self-Depreciation	.09	20	.14	20
Inappropriate Behaviour	-.34	21	-.46	21
Inferences	-.52	16	-.67	16

Notes

Truth=1, Lie= -1, a positive correlation indicates the item indicated veracity of interviewees or evaluators' propensity to make a truth judgement

Correlations above .40 are highlighted in bold, as recommended by Fergusson (2009)

Discussion

Akehurst, et al., (2015) reported that a checklist derived from the literature on lying and malingering assisted interviewers in the correct classification both of truthful interviewees as well as those exaggerating their symptoms. Analysis of results for the 28-item checklist identified 9 items which significantly discriminated between truth tellers and exaggerators. The current study sought to investigate the utility of an abbreviated version of the full checklist based on those 9 items. Evaluators using that abbreviated version of the original checklist did not perform at a level significantly above that expected by chance, and their performance was no better than a control group who were not given any checklist (64% accuracy for both conditions). Evaluators in the current study who used the full checklist, however, achieved 86% accuracy (100% for truth-tellers and 73% for exaggerators), providing further support for previous findings which highlighted the apparent utility of the checklist (Akehurst et al., 2015), wherein interviewers using the checklist achieved 75% correct classification of truthful interviewees and 66% correct classification of those who were exaggerating their symptoms.

These positive results for the full checklist are similar to those reported in laboratory studies for investigative interviewers who have been trained to use CBCA and/or RM techniques, wherein success rates have ranged from 65% to 80% (Vrij, 2008), and so these findings would appear to lend support to the proposal that the detection of lying or malingering in the medico-legal interview setting might be improved given use of an appropriate checklist.

Whereas, in the first study based on use of the checklist, nine checklist items were found to significantly discriminate between truth tellers and exaggerators, the whole appears to be greater than the sum of its parts. Moreover, when comparing the utility of items with the Akehurst et al. 2015 findings, it was a different sub-sample of items in the current study that were most effective at discriminating truths from exaggeration. Table 1 illustrates that (i) attribution of recklessness, (ii) reported discomfort, (iii) quantity of detail, (iv) temporal details, (v) long latency periods, and (vi) inferences were most effective in the current study. It is likely that the *context* of the full checklist may help structure understanding of the items. Detailed analysis of the findings indicated that evaluators were better able to use some items of the checklist (i.e. responses to some items were predicting the intended quality of the target) than others.

Evaluators appeared to have different readings or interpretations of some items, thus leading to relatively large variability in the coding of some of the items. This discrepancy may be exaggerated in a reduced checklist, as the codings made by the full checklist group were typically the best predictors of actual veracity. Further research might usefully explore the impact on accuracy of judgements (positive or negative) of training and/or repeated use of the checklist.

The assessment of truth telling and endeavour to deceive is undertaken in the medico-legal setting where estimates of base rates vary widely, and may differ between assessed groups. Thus, while Ardolf, Denney, and Houston (2007) have estimated that probable and definite malingered neurocognitive dysfunction in a sample of criminal defendants was 54.3%, estimated rates of malingered symptoms after personal injury have ranged from 1% to 50% (Appelbaum et al., 1993). Detection of deception is, therefore, perhaps more of an art than a science. The checklist seeks to

assist the expert and inform the development of opinion, and requires interpretation by the expert in the broader context of the interview and in the light of other information. Such a checklist seeks to draw the expert's attention to certain aspects of the interview, and it does appear that it is not simply the pause for reflection that assists in the process of distinguishing between truthful and false reporting. In the busy and increasingly time pressured medico-legal assessment setting, benefits associated with the use of a brief aide-memoire might be associated with an enhancement of the expert assessor's ability to assess credibility. The appropriate use of such an aide-memoire may also serve to reassure the court, and those relying on an expert report, that the assessor has actively considered relevant applied research findings. Experts may not otherwise draw upon the relevant research if they are unfamiliar with it, or they may not take the time to more formally or explicitly consider an interview in the light of the relevant literature.

Participants were motivated to malingering in the study with a suggestion that, if they convinced the interviewer of their credibility, they might imagine being awarded compensation. However, this does not mirror the motivation of interviewees in a medico-legal setting in the real world. Further exploration of the degree to which participants were motivated to exaggerate convincingly during their interviews is warranted, and the influence of higher degree of incentive to lie successfully in particular would be worthy of attention.

When considered alongside the previous study (Akehurst, et al., 2015) of the checklist, further exploration of the device and its potential use in the applied medico-legal setting appears warranted, so as to explore generalizability of the findings.

Informal feedback indicates that a brief aide-memoire may be acceptable in the medico-legal assessment setting. Expert witnesses may value the opportunity to consider aspects from research with which they may not be familiar, although it may be the case that it will be useful to explore the utility of combining the device with other assessments used in various specialist settings. Thus, further research would usefully look into the use of this device alongside neuropsychological tests of malingering or Symptom Validity Testing in medically unexplained symptoms, for example, noting that there are concerns about the validity of such tests, and some have expressed concerns about using too many assessments (Leighton, Weinborn, & Maybery, 2014; Merten, & Merkelbach, 2013; Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013). If an aide-memoire of the kind presented here is to be considered for use in the medico-legal setting, where both experts and the court need to be confident in methods and procedures used in the generation of opinions, great care needs to be taken to ensure that benefits of adoption outweigh costs and risks. It would seem appropriate to explore the relevance and impact of training in the use and interpretation of such a device.

A checklist is only as good as its interpretation, and misinterpretation can easily negate any potential benefit.

Conclusion

This study indicates that the utility of an aide-memoire in the form of a 28-item checklist derived from the literature on lying and malingering (Akehurst et al., 2015) is adversely affected by item reduction, and provides further support to the indication that the full checklist improves assessors'

ability to discriminate between truth tellers and exaggerators. The use of such a checklist in the medico-legal assessment setting could assist expert witnesses as they form the opinion expressed, and offer the Courts some reassurance that the expert assessor has considered more relevant aspects of an interviewee's presentation. However, undue reliance on the checklist results as if a formal psychometric test is to be actively avoided, noting criticism in some quarters of more formal use of measures such as the Structured Inventory of Reported Symptoms (Rogers, Bagby, & Dickens, 1992; Rogers et al., 2010) and in the court setting (DeClue, 2011). Further research should explore requirement for training in the use of the device, assessment of its utility in the applied medico-legal setting, and the costs and benefits of its combination with other assessments. Our findings complement the findings of Akehurst et al., (2015), and lend weight to the proposal that the 28-item checklist may hold considerable value as an efficient structured decision aid for time-pressured expert witnesses in the medico-legal setting.

References

- Akehurst, L., Easton, S., Fuller, E., Drane, G., Kuzmin, K., & Litchfield, S. (2015). An evaluation of a new tool to aid judgements of credibility in the medico-legal setting. *Legal and Criminological Psychology*. doi: 10.1111/lcrp.12079
- Appelbaum, P. S., Jick, R. Z., Grisso, T., Givelber, D., Silver, E., & Steadman, H. J. (1993). Use of posttraumatic stress disorder to support an insanity defense. *The American Journal of Psychiatry*, 150(2), 229. Retrieved from <http://europepmc.org/abstract/MED/8422072/reload=0;jsessionid=l2h4Z6WTifluElx44IVu.16>
- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, 21, 899-916. doi: 10.1080/13825580600966391.
- Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury*, 27(7-8), 909-916. doi:10.3109/02699052.2013.793400
- Blandón-Gitlin, I., Pezdek, K., Lindsay, D.S., & Hagen, L. (2009). Criteria-based Content Analysis of true and suggested accounts of events. *Applied Cognitive Psychology*, 23, 901–917. doi: 10.1002/acp.1504.
- Bond, C.F. Jr., & DePaulo, B.M. (2006). Accuracy of deceptive judgements. *Personality and Social Psychology Review*, 10, 214 – 234. doi: 10.1207/s115327957pspr1003_2.
- DeClue, G. (2011). Harry Potter and the structured interview of reported symptoms? *Open Access Journal of Forensic Psychology*, 3, 1-17. Retrieved from http://www.forensicpsychologyunbound.ws/OAJFP/Forensic_Instruments_files/DeClue%202011.pdf
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, 129(1), 74. Retrieved from <http://dx.doi.org/10.1037/0033-2909.129.1.74>
- Drob, S. S., Meehan, K. B., & Waxman, S. (2009). Clinical and conceptual problems in the attribution of malingering in forensic evaluations. *Journal of the American Academy of Psychiatry and Law* 2009; 37: 98 – 106. Retrieved from <http://www.jaapl.org/content/37/1/98.full.pdf+html>
- Edens, J., Otto, R., Buffington, J., Tomicic, T., & Poythress, N. (2001) Factors differentiating successful versus unsuccessful malingerers. *Journal of Personality Assessment*, 77, 333-338. doi: 10.1207/S15327752JPA7702_13.

- Faust, D., & Ahern, D. (2012). Clinical Judgement and Prediction, chapter 8 in Faust, D. *Coping with psychological and psychiatric testimony*. 6th Ed. OUP. New York.
- Faust, D., Ahern, D., Bridges, A., & Yonce, L. (2012). Assessment of Malingering and Falsification: Pushing the Boundaries of Knowledge in Research and Clinical Practice; chapter 2 in Reynolds, C. R., & Horton, A. M. (eds.), *Detection of Malingering during Head Injury Litigation*, Springer Science + Business Media, LLC
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538. doi:10.1037/a0015808
- Ferrari, R., Kwan, O., Russell, A., Pearce, J., & Schrader, H. (1999). The best approach to the problem of whiplash? One ticket to Lithuania please. *Clinical and Experimental Rheumatology*, 17, 321-326. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10410265>
- Hall, R., & Hall, R. (2006). Malingering of PTSD: forensic and diagnostic considerations, characteristics of malingerers and clinical presentations. *General Hospital Psychiatry*, 28 (6), 525-535. doi: 10.1016/j.genhosppsy.2006.08.011.
- Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 137(4), 643–659. doi:10.1037/a0023589
- Hinkin, T. R. (1995). A review of scale development practises in the study of organisations. *Journal of Management*, 21(5): 967-988.
- Johnson, M. K., and Raye, C. L. (1981). Reality Monitoring. *Psychological Review*, 88, 67-85. doi: <http://doi.apa.org/index.cfm?fa=search.displayRecord&uid=1981-06694-001>.
- Kramer, G. M., & Gagliardi, G. J. (2009). Forensic evaluation of insanity: Assessing valid symptom report in defendants with major mental disorder. *Journal of Forensic Psychology Practice*, 9(1), 92 – 102. doi: 10.1080/15228930802427171.
- Leighton, A., Weinborn, M., & Maybery, M. (2014). Bridging the gap between neurocognitive processing theory and performance validity assessment among the cognitively impaired: A review and methodological approach. *Journal of the International Neuropsychological Society*, 20(09), 873-886. doi.: <http://dx.doi.org/10.1017/S135561771400085X>
- Maguire, B., Harvey, A., & Shores, E. (2001). Simulated malingering in pain patients: a study with the pain patient profile. *British Journal of Clinical Psychology*, 40, 71-79. doi: 10.1348/014466501163490.
- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime and Law*, 11, 99–122. doi:10.1080/10683160410001726356.

- Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law*, 6(2), 122-137. doi: 10.1007/s12207-013-9155-x
- Miller, H. A. (2001). *Miller-Forensic Assessment of Symptoms Test (M-FAST): Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Miller, H. A. (2004). Examining the use of the M-FAST with criminal defendants incompetent to stand trial. *International Journal of Offender Therapy and Comparative Criminology*, 48(3), 268-280.
- Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, 24(6), 643. Retrieved from <http://dx.doi.org/10.1023/a:1005500219657>
- Resnick, P. J. (1995). Guidelines for the evaluation of malingering in posttraumatic stress disorder. In R. I. Simon (Ed.), *Posttraumatic stress disorder in litigation: Guidelines for forensic assessment* (pp. 117-134). Washington: American Psychiatric Press.
- Resnick, P. J. (1997). Malingered psychosis. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed.) (pp. 47-67). New York: Guilford Press.
- Resnick, P. J., & Knoll, J. L. (2005). Faking it: How to detect malingered psychosis. *Current Psychiatry*, 4, 13-25.
- Rogers, R. (1997). *Clinical assessment of malingering deception*. New York: Guilford.
- Rogers, R. (2008). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception*, (3rd ed., 14-38). New York, NY: Guilford.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS)*, Professional manual. Lutz, FL: Psychological Assessment Resources, Inc.
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms (SIRS)*, 2nd Edition, professional manual. Lutz, FL: Psychological Assessment Resources, Inc.
- Sporer, S. (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. doi: 10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0.

- Strömwall L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology, 18*, 653 – 668. doi: 10.1002/acp.1021.
- Vrij, A. (2008). *Detecting lies and deceit: pitfalls and opportunities*. Chichester: Wiley.
- Vrij, A., Akehurst, L., Brown, L., & Mann, S. (2006). Detecting lies in young children, adolescents and adults. *Applied Cognitive Psychology, 20*, 1225 – 1237. doi: 10.1002/acp.1278
- Vrij, A., Akehurst, L., & Knight, S. (2006). Police officers', social workers', teachers' and general public's beliefs about deception in children, adolescents and adults. *Legal and Criminological Psychology, 11*, 297 – 312. doi: 10.1348/135532505X60816.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004a). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research, 30*(1), 8 – 41. doi: 10.1111/j.1468-2958.2004.tb00723.x.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004b). Let me inform you about how to tell a convincing story: CBCA and Reality Monitoring scores as a function of age, coaching and deception. *Canadian Journal of Behavioural Science, 36*, 113 – 126. doi: 10.1037/h0087222.
- Vrij, A., Edward, K., & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin, 27*, 899-909. doi: 10.1177/0146167201277012.
- 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. doi: 10.1177/1529100610390861.
- Zapf, P. & Grisso, T. (2012). Use and misuse of forensic assessment instruments; chapter 19 in Faust, D. *Coping with psychological and psychiatric testimony*. 6th Ed. OUP. New York.

Appendix 1: The checklist items and their sources

Items in full checklist	Item source	Item included in abbreviated checklist
If the individual pauses when describing symptoms, do the pauses last longer than when answering other questions?	DePaulo et al. (2003) Vrij (2008)	
Does the individual show long latency periods, i.e. the time between the end of a question and the beginning of an answer?	DePaulo et al. (2003) Vrij (2008)	
Does the individual, when talking about the cold plunge test and subsequent experience, talk in a disorganised way rather than speak in a structured and chronological order?	CBCA criterion 2	
Does the individual, when talking about the cold plunge test and subsequent experience, give substantial quantity of detail?	CBCA criterion 3	
Does the individual, when talking about the cold plunge test and subsequent experience, describe unusual detail and/or unexpected complications?	CBCA criteria 7 & 8	Yes
Does the individual, when talking about the cold plunge test and subsequent experience, provide superfluous details (e.g. description of details which are not really relevant to answering the questions)?	CBCA criterion 9	
Does the individual, when talking about the cold plunge test and subsequent experience, use terms and language or show knowledge related to his/her symptoms that he/she would not be expected to have?	CBCA criterion 10	
Does the individual, when talking about the cold plunge test and subsequent experience, provide accounts of his/her own mental state?	CBCA criterion 12	Yes
Does the individual, when talking about the cold plunge test and subsequent experience, provide <i>spontaneous corrections</i> ?	CBCA criterion 14	
Does the individual, when talking about the cold plunge test and subsequent experience, tend to <i>admit to a lack of memory</i> ?	CBCA criterion 15	
Does the individual, when talking about the cold plunge test and subsequent experience, <i>raise doubts about his/her own account</i> ?	CBCA criterion 16	
Does the individual, when talking about the cold plunge test and subsequent experience, show <i>self-deprecation (i.e. put him or herself down)</i> ?	CBCA criterion 17	
Does the individual show <i>inappropriate behaviour</i> ?	Validity Checklist item 2	
Does the individual give clear answers to questions about the symptoms?	Reality Monitoring criterion 1	

Does the individual provide detail about <i>physical sensations</i> (e.g. "It felt like pins and needles")?	Reality Monitoring criterion 2	
Does the individual provide <i>temporal details</i> ? (e.g. details about time order of the events; "the pins and needles came before the numb feeling")	Reality Monitoring criterion 4	
Does the individual provide details about <i>duration of events</i> ? ("the numbness lasted for about 10 minutes.")	Reality Monitoring criterion 4	
Does the individual give plausible answers to the questions about the symptoms?	Reality Monitoring criterion 7	
Does the individual show cognitive operations: <i>inferences at the time of the event or at interview</i> ? (e.g. it appeared to me that she was enjoying me doing the ice plunge")	Reality Monitoring criterion 8	
Is the individual falsely ascribing real symptoms to the cold plunge test (e.g. they had dry skin anyway and now blaming the cold plunge)?	Resnick (1997)	
Does the individual show lack of co-operation at interview?	Resnick (1997)	
Does the individual blame someone else for his/her symptoms?	Ferrari et al. (1999)	Yes
Does the individual believe that the other person behaved recklessly, so causing their discomfort?	Ferrari et al. (1999)	Yes
Does the individual expect his/her discomfort to last a short time and/or show significant improvement?	Ferrari et al. (1999)	Yes
Does the individual tend to confirm existence of many symptoms?	Rogers (1997)	Yes
Are there symptoms which do not readily fit those expected with a cold plunge test?	Rogers (1997) Validity Checklist item 1	Yes
Are there unlikely or contradictory patterns of symptoms?	Rogers (1997) Validity Checklist item 1	Yes
Are some symptoms reported as at extreme, improbable levels of severity?	Rogers (1997)	Yes