SHORT PAPER

WILEY

Effects of time pressure on strategy selection and strategy execution in forced choice tests

Robin Orthey[1,2] | Nicola Palena[3] | Aldert Vrij[1] | Ewout Meijer[2] | Sharon Leal[1] | Hartmut Blank[1] | Letizia Caso[3]

[1] Department of Psychology, University of Portsmouth, Portsmouth, UK

[2] Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

[3] Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy

Correspondence
Robin Orthey, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.
Email: robinorthey@googlemail.com

Summary

We examined the effects of cognitive load on the strategy selection in the forced choice test (FCT) when used to detect hidden crime knowledge. Examinees ($N = 120$) with and without concealed knowledge from a mock crime were subjected to an FCT either under standard circumstances or cognitive load. Cognitive load was implemented through time pressure. The FCT distinguished examinees with concealed knowledge from those without better than chance in both conditions, but the counter-strategies did not differ between conditions. Further investigation revealed that time pressure did affect examinees' ability to follow their intended counterstrategy to produce randomized test patterns, which constitutes an effective counterstrategy in the FCT. Hence, time pressure lowered the success rate of effective counterstrategies, but not their incident rates. Further disambiguation of various cognitive load manipulations and their effects on strategy selection and execution is needed.

KEYWORDS
2AFCT, cognitive load, concealed knowledge detection, FCT, time pressure

1 | INTRODUCTION

The forced choice test (FCT) can be applied to detect concealed knowledge about an event (Denney, 1996; Pankratz, 1983). In an FCT, the examinees are presented with questions about the event, two possible answer alternatives, and the instruction to select the correct answer alternatives or to guess in case they do not know. Although examinees who are unaware of the correct answer have no choice other than to guess, examinees who try to conceal their knowledge tend to purposefully select incorrect answers. Therefore, test scores fall below chance levels—so-called underperformance—and can be used as detection criterion (Bianchini, Mathias, & Greve, 2001; Van Oorsouw & Merckelbach, 2010).

Empirical research suggests that examinees with concealed knowledge can successfully be detected at rates varying from 40% to 60% while maintaining a low false positive rate at around 5% (Giger, Merten, Merckelbach, & Oswald, 2010; Jelicic, Merckelbach, & van Bergen, 2004; Meijer, Smulders, Johnston, & Merckelbach, 2007; Merckelbach, Hauer, & Rassin, 2002; Orthey, Vrij, Leal, & Blank, 2017; Shaw, Vrij, Mann, Leal, & Hillman, 2014). This detection accuracy is directly related to the prevalence of three different self-reported response patterns that examinees with concealed knowledge use to avoid being detected by the test (Orthey et al., 2017; Orthey, Vrij, Meijer, Leal, & Blank, 2018). These response patterns are defined in terms of hierarchical strategy levels and specify how answer alternatives are selected depending on the examinees' beliefs about the test's detection mechanism (Orthey et al., 2017; Orthey et al., 2018). Examinees using a level 0 strategy form no belief about the test's detection mechanism and comply with the test instructions to select the correct answer alternatives. Examinees using a level 1 strategy assume that the

test's detection mechanism is based on a level 0 strategy and their response pattern is a reaction to the test instructions. Instead of selecting the correct answers, examinees select the incorrect answers. Examinees using a level 2 strategy assume that the test uses a level 1 strategy as detection mechanism and provide a mixture of correct and incorrect answers as response pattern instead. Although each strategy level predicts a different behaviour, the intended objective is the same, namely, to avoid detection by the FCT. In an FCT, levels 1 and 2 are the most prevalent strategy levels with roughly equal frequencies; level 0 rarely occurs in examinees with concealed knowledge. Consequently, the underperformance criterion used to detect concealed knowledge in an FCT is suitable for detecting a level 1 strategy but does not detect examinees using a level 2 strategy. Therefore, in theory, detection accuracy could be increased by manipulations that shift the participant's strategy from level 2 to level 1.

The three strategy levels were derived from cognitive hierarchy theory (see Carmerer, Ho, & Chong, 2004). From this theory, it follows that limitations in cognitive resources affect the strategy selection. As such, the strategy an examinee selects is not necessarily the optimal strategy but rather a strategy that is "good enough" given the available cognitive resources (also known as satisficing; see Simon, 1955). Previous research indicates that a large proportion of examinees have sufficient cognitive resources available to discern the test's mechanism and to devise an appropriate counterstrategy (see Orthey et al., 2017; Orthey et al., 2018). Thus, if one could limit the cognitive resources available to examinees, this would reduce the frequency of higher order strategies (e.g., level 2). As a consequence, the detection accuracy of the FCT would increase, because more examinees would be forced to employ a level 1 strategy instead.

It is generally accepted that humans have a limited amount of cognitive resources available at any given moment. Therefore, increasing the overall cognitive load limits the available resources that can be allocated (Plass, Moreno, & Brunken, 2010). We chose to implement cognitive load through time pressure, as it is a commonly used manipulation for cognitive load (see Klapproth, 2008) and it can easily be introduced into the FCT paradigm. Hence, we subjected examinees to a mock crime procedure or a filler task followed by either a standard FCT or an FCT with the restriction that each question had to be answered within 2 s. We tested two hypotheses: Under time pressure, examinees will be more likely to report using lower level strategies (e.g., level 1 instead of level 2, or level 0 instead of level 1) than under standard conditions (Hypothesis 1). As a consequence, examinees with concealed knowledge will display more extreme (positive or negative) test scores, resulting in increased classification accuracy of the FCT (Hypothesis 2).

2 | METHOD

2.1 | Participants

We tested 120 examinees (33 males and 87 females) from a university undergraduate population. Their mean age was $M = 24.61$ ($SD = 7.31$).

The experiment was approved by the ethics committee of the Science Faculty of the University of Portsmouth, and examinees received course credit for participating in the experiment.

2.2 | Procedure

Examinees were randomly assigned to one of two virtual reality scenarios. In both scenarios, examinees were placed in a virtual apartment that could be freely explored from the first-person perspective. In the concealed knowledge conditions, examinees were told that they were to investigate the apartment of a terrorist and had to obtain and remember as much information as possible about the terrorist and his planned actions. The apartment contained clues that could be investigated further. These clues were easily visible. Examinees could examine them further by clicking on them. This provided them with a more detailed picture and short description of the clue. Once all clues were examined, the simulation terminated and examinees were instructed not to reveal the knowledge gained from the simulation for the remainder of the experiment.[1] In the no concealed knowledge condition, examinees were instructed to survey a different apartment and instructed to remember as much detail as possible. This simulation terminated after 3 minutes.

Then all examinees were subjected to an FCT examination about the terrorist apartment. The test was computerized, and examinees were randomly assigned to either the standard or time pressure condition. In the standard conditions, examinees received 20 questions about the terrorist apartment, each of which featured two possible answer alternatives. Questions were presented in two steps. First, the question was displayed in the centre of the screen. Upon clicking the "next" button at the bottom centre of the screen, the question disappeared and the two answer alternatives were presented at the top left/right side. All answer alternatives were pictures, and examinees could select an answer by clicking on it with a mouse button. Examinees received the following instructions: "Next, you will be presented with a number of questions and two answer alternatives per question. Select the correct answer. If you don't know, guess." Examinees in the time pressure condition received the additional instruction: "You have to choose an answer alternative for each question within two seconds, otherwise the trial will time-out. If you time-out too often, you fail the test automatically." In case an examinee took longer than 2 s, a buzzer sound occurred to signal the time out. The number of time outs had no consequences for the rest of the experiment.

After the FCT procedure, examinees were instructed that the deception detection task was over and that they should answer the following questions honestly. To measure the response strategies used during the FCT, examinees were asked: "What did you do to avoid being classified as a liar by the previous test?" Their answers were recorded, transcribed, and coded by two independent coders (90.00% absolute agreement).

---

[1]Because of restriction imposed by our ethical committee, we were not allowed to ask participants to imagine that they were terrorists. We therefore chose this compromise.

Finally, examinees with concealed knowledge received the 20 FCT questions and answer alternatives again and were tasked to indicate the correct answer alternative they remembered from the simulation. This served as a memory check, and memory performance was good (91.17%).

2.3 | Materials

We used the same virtual reality simulations, FCT questions, and answer alternatives as in Orthey et al. (2018). The answer alternatives of all questions were validated to be equally plausible (see Doob & Kirschenbaum, 1973). Orthey et al. (2018) validated the questions and answer alternatives by asking examinees who were unaware of the correct answers to indicate the most plausible answer. Any questions in which one answer alternative was selected by 70% or more of the sample were deemed biased and removed. In total, the FCT contained 20 questions with two answer alternatives each. Answer alternatives were presented pictorially and had the same size. To control for order effects, the sequence of questions was counterbalanced across examinees, using a Latin square of the size 20. Therefore, the 20 questions occurred equally often over all possible trials (1–20). The horizontal alignment of the correct answer alternative was determined randomly on each trial.

2.4 | Design

This experiment featured a 2 Veracity (concealed knowledge vs. no concealed knowledge) × 2 Cognitive load (standard vs. time pressure) between-subjects design with the test scores as dependent variable. Test scores were computed by submitting the raw total number of correct answer alternatives selected to a z transformation according to the binomial distribution (see Siegel & Castellan, 1988, p. 43). The higher/lower a z score was, the less likely it was to occur due to chance. Underperformance was defined as test scores below chance performance. Detection accuracy was estimated using the area under the curve (AUC) of the receiver operating characteristic (see Tanner & Swets, 1954; Hanley & McNeil, 1982). A receiver operating characteristic plots the sensitivity (detection rate of examinees with concealed knowledge) against the false positive rate (incorrect classifications of examinees without concealed knowledge) for each possible cut-off. The AUC indicates the general diagnostic value for all possible cut-offs. An AUC ranges from 0 to 1, and 0.5 refers to chance performance. AUCs larger than 0.5 suggest that the criterion detects the signal better than chance. Examinees' test scores were used to postdict examinees' true group membership (concealed knowledge/no concealed knowledge). Lower scores were treated as indicative of concealed knowledge. For comparison with previous articles, we also report sensitivity and specificity (detection rate of examinees with and without concealed knowledge). For this analysis, examinees with a z score smaller than −1.65 were classified as having concealed information (see also Meijer et al., 2007; Verschuere, Meijer, & Crombez, 2008). For further exploration, we categorized FCT z scores into below, above, and within chance performance as predicted by the three strategy levels. We used the traditional 5% cut-off (bidirectional; z scores larger than 1.96 for above chance performance, or lower than −1.96 for below chance performance) for classification.

Examinees' responses to the open-ended question were coded into distinct strategy levels (0, 1, and 2; see Orthey et al., 2017; Orthey et al., 2018). These strategy levels define the self-reported selection strategy of the examinee based on their belief over the tests' detection mechanism. A level 0 strategy assumes that no belief over the detection mechanism are formed and examinees comply with the test instructions to select the correct answer alternatives. Hence, a level 0 strategy would result in performance above chance. A level 1 strategy operates on the belief that the test identifies concealed knowledge through compliance with the test instructions and therefore, features a reaction to them, such as picking the incorrect answers instead. Employing a level 1 strategy would result in performance below chance. Finally, a level 2 strategy is derived from the understanding that the test detects concealed knowledge through underperformance. Consequently, a level 2 strategy features behaviours with the goal to provide a mixture of correct and incorrect answers, resulting in test scores that fall within chance performance.

3 | RESULTS

3.1 | Reaction time

To test whether our time pressure manipulation was effective, we conducted a Welch's t test for the time pressure condition with average reaction time as dependent variable. We found a significant difference between the standard and time pressure conditions, $t(63.93) = 9.38$, $p < .001$. Examinees in the time pressure condition ($M = 1.48s$, $SD = 0.81$) responded faster than examinees in the standard condition ($M = 6.43s$, $SD = 3.99$).[2] Furthermore, we looked into the number of times participants timed out (i.e., a response time longer than 2 s). In the time pressure condition, examinees with and without concealed knowledge differed significantly in the number of times the time limit was exceeded, $t(58) = 3.59$, $p < .001$, examinees with concealed knowledge timed out on average 1.03 ($SD = 0.99$) times, and examinees without concealed knowledge timed out on average 4.07 ($SD = 4.52$) times. In sum, we conclude that our time pressure

---

[2]We also differentiated for the factor veracity. A 2 Veracity (concealed knowledge vs. no concealed knowledge) × 2 Cognitive load (standard vs. time pressure) between-subjects analysis of variance with average reaction time as dependent variable was conducted. We found a significant difference for Veracity, $F(1, 116) = 18.91$, $p < .001$, $\eta^2 = .14$. Examinees without concealed knowledge ($M = 4.99$, $SD = 4.58$) took longer than those with concealed knowledge ($M = 2.92$, $SD = 2.43$). We also found a significant effect for cognitive load, $F(1, 116) = 108.92$, $p < .001$, $\eta^2 = .48$. Examinees in the standard condition ($M = 6.43$, $SD = 3.99$) took longer than examinees in the time pressure condition ($M = 1.48$, $SD = 0.81$). There was also a significant Veracity × Cognitive load interaction, $F(1, 116) = 11.11$, $p = .001$, $\eta^2 = .09$. The difference between examinees with and without concealed knowledge was larger in the standard condition (concealed knowledge: $M = 4.61$, $SD = 2.47$; no concealed knowledge: $M = 8.25$, $SD = 4.42$) than in the time pressure condition (concealed knowledge: $M = 1.25$, $SD = 0.22$; no concealed knowledge: $M = 1.73$, $SD = 1.09$), $F(1, 116) = 29.51$, $p < .001$.

manipulation resulted in faster responses and a low frequency of time outs.

3.2 | Response strategies

First, we examined the self-reported strategy levels of examinees with concealed knowledge. In both the standard and time pressure conditions, level 1 strategies were the most prevalent (standard = 48%; time pressure = 62%) followed by level 2 strategies (standard = 33%; time pressure = 28%) and level 0 strategies (standard = 19%; time pressure = 10%). A chi-square test of independence was calculated comparing the frequency of the strategy levels between the standard and time pressure conditions. We found that the frequency of the different strategy levels did not differ between conditions, $X^2(2, N = 56) = 1.30, p = .523$. These results do not support Hypothesis 1.

3.3 | Test scores

The test scores detected concealed knowledge better than chance in both the standard and time pressure conditions (see Table 1). To further assess the effects of time pressure on the test scores, we compared examinees with concealed knowledge between conditions per strategy levels. First, we categorized the test scores into performance below, above, and within chance levels. Table 2 displays frequency of test outcomes per condition and within each self-reported strategy level. For level 0 and 1 strategies, the distributions of test scores were similar. For the examinees using level 2 strategies, in the standard condition, 89% fell within chance performance and only 11% fell below chance performance. In the time pressure condition, only 37.5% fell within chance performance with 50% displaying below chance level performance. Time pressure seemed to affect only examinees who reported to randomize between correct and incorrect answers, so we tested whether scores outside chance performance (below and above chance performance combined) were more likely to occur under time pressure for level 2 strategies. A chi-square test revealed a significant effect, $X^2(1) = 4.90, p = .027$; test scores outside chance performance occurred more frequently under time pressure. Additionally, we conducted an independent samples t test on the absolute test scores, because not all assumptions of the chi-square test were met. Examinees who reported a level 2 strategy had higher test scores in the time pressure condition ($M = 1.90, SD = 0.62$), than

TABLE 1 Detection accuracy of total scores per condition

Condition	Sensitivity (%)	Specificity (%)	AUC	p	95% CI
Test scores					
Standard	50.00	97.77	0.66	.034	[.50, .82]
Time pressure	66.66	93.33	0.80	<.001	[.67, .93]

Note. Sensitivity and specificity were calculated using the traditional 5% cut-off ($z < −1.65$). AUC denotes area under the curve, and 95% confidence intervals (CIs) are reported.

TABLE 2 Examinees' total scores categorized into below, above, and within chance performance

Condition	Below chance performance	Within chance performance	Above chance performance	N
Level 0				
Standard	1	0	4	5
Time pressure	1	0	2	3
Level 1				
Standard	13	0	0	13
Time pressure	14	2	2	18
Level 2				
Standard	1	8	0	9
Time pressure	4	3	1	8

Note. Frequencies are differentiated per strategy level for examinees with concealed knowledge. Scores were categorized as follows: below chance performance: $x \leq −1.96$; above chance performance: $x \geq 1.96$; and within chance performance: $−1.96 < x < 1.96$. Examinees that could not be categorized in any of these strategy levels were excluded. This led to three exclusions in the standard condition and one in the time pressure condition.

in the standard condition ($M = 0.77, SD = 0.69$), $t(15) = −3.50, p = .003$. This effect was strong (Cohen's $d = 1.72$). Altogether, this supports our second hypothesis that test scores become more extreme under time pressure although only for examinees who reported to use a level 2 strategy.

4 | DISCUSSION

We subjected examinees with and without concealed knowledge to a standard FCT or a modified FCT that forced examinees to respond within 2 s for each question. We introduced time pressure to the FCT paradigm to elicit cognitive load, which we expected would lead examinees with concealed knowledge to be more likely to select lower level strategies and hence, increase the detection accuracy of the FCT.

We found no evidence that time pressure affects strategy selection. Although participants in the time pressure condition responded faster, examinees did not differ in the frequencies of self-reported response strategies, and these frequencies matched those found in other experiments (see Orthey et al., 2017; Orthey et al., 2018). Yet further exploration suggests that time pressure did lead to significantly more extreme test scores in examinees with concealed knowledge who reported using level 2 strategies. When categorized into below, above, and within chance performance, more than half of those examinees fell outside the chance performance category, and only a minority managed to achieve chance performance (around 37%). This stands in sharp contrast with findings in our control condition and previous research (see Orthey et al., 2017; Orthey et al., 2018), where most examinees who reported to have randomized their answers achieved test scores that fall within chance performance.

Thus, even though time pressure did not affect the strategy examinees with concealed knowledge reported to have used, it did affect their ability to successfully execute these strategies. These findings have important implications for increasing the detection accuracy of the FCT. Time pressure is easy to implement into an FCT procedure and could be especially useful when the prevalence of randomization behaviour is high. This could, for example, be the case when examinees have informed themselves about the rationale underlying the test by reading articles, coached by an attorney, or after repeated testing (for the effects of coaching on the FCT, see Verschuere et al., 2008).

In terms of overall detection accuracy, both the traditional FCT and the time pressure FCT detected concealed knowledge better than chance. The standard condition had a detection accuracy close to AUC = 0.70, which is within the range of previous research (Meijer et al., 2007; Orthey et al., 2017; Orthey et al., 2018). By comparison, the time pressure condition featured one of the best detection accuracies found so far, around AUC = 0.80. A likely reason for this is the reduced success rate of level 2 strategies, resulting in more extreme test scores that are detected by the underperformance criterion. This implies that detection accuracy could additionally be increased by making effective counterstrategies harder to perform successfully.

From a theoretical point of view, these findings suggest that strategy selection is not the only component affecting the test score. In addition, examinees' ability to successfully execute their intended strategy plays a role. In this case, time pressure led examinees who reported using a level 2 strategy to produce more extreme test scores than those not under time pressure. Other, lower level strategies were not affected, likely because they are easier to execute (i.e., either selecting only correct or only incorrect answers). That means the influence of cognitive load must be differentiated in terms of affecting strategy selection or strategy execution. Further disambiguation between strategy selection, the intended test outcome, and strategy execution, the actual test outcome, is needed, especially in light of various implementations of cognitive load.

Future research on the strategy selection could focus more on making it harder to discern the test's detection mechanism through misdirection (see Kuhn, Caffaratti, Teszka, & Rensink, 2014). For example, Orthey et al. (2017) added a fake polygraph to the set-up of the FCT procedure in order to make examinees believe their physiological responses were recorded during the test. As a consequence, more examinees complied with the test instructions to select the correct answer alternatives (lowest level strategy) with the polygraph set-up than in the control group. In a similar manner, other, more salient forms of misdirection could be used to shape examinees' strategy selection process.

In sum, although we found no evidence that time pressure affects the strategy selection of examinees with concealed knowledge, it did affect the execution of their self-reported strategy, resulting in lower success rates of level 2 strategies. As such, time pressure provides an easy to implement adjustment to the FCT that will likely increase its detection accuracy.

CONFLICT OF INTERESTS

There are no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data is available for download at: https://osf.io/6dr52/?view_only=cda1cf771c604df6b828d94245808d97

ORCID

Robin Orthey https://orcid.org/0000-0002-6185-3061
Nicola Palena https://orcid.org/0000-0002-2746-1208
Aldert Vrij https://orcid.org/0000-0001-8647-7763

REFERENCES

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist*, 15(1), 19–45. https://doi.org/10.1076/clin.15.1.19.1907

Carmerer, C. F., Ho, T., & Chong, J. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119, 861–898. https://doi.org/10.1162/0033553041502225

Denney, R. L. (1996). Symptom validity testing of remote memory in a criminal forensic setting. *Archives of Clinical Neuropsychology*, 11(7), 589–603. https://doi.org/10.1093/arclin/11.7.589

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups—Partial remembering. *Journal of Police Science and Administration*, 1, 287–293.

Giger, P., Merten, T., Merckelbach, H., & Oswald, M. (2010). Detection of feigned crime-related amnesia: A multi-method approach. *Journal of Forensic Psychology Practice*, 10, 440–463. https://doi.org/10.1080/15228932.2010.489875

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Jelicic, M., Merckelbach, H., & van Bergen, S. (2004). Symptom validity testing of feigned amnesia for a mock crime. *Archives of Clinical Neuropsychology*, 19, 525–531. https://doi.org/10.1016/j.acn.2003.07.004

Klapproth, F. (2008). Time and decision making in humans. *Cognitive, Affective, & Behavioural Neuroscience*, 8, 509–524. https://doi.org/10.3758/CABN.8.4.509

Kuhn, G., Caffaratti, H. A., Teszka, R., & Rensink, R. A. (2014). A psychologically-based taxonomy of misdirection. *Frontiers in Psychology*, 5, 1–14. https://doi.org/10.3389/fpsyg.2014.01392

Meijer, E. H., Smulders, F. T., Johnston, J. E., & Merckelbach, H. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology*, 44, 814–822. https://doi.org/10.1111/j.1469-8986.2007.00543.x

Merckelbach, H., Hauer, B., & Rassin, E. (2002). Symptom validity testing of feigned dissociative amnesia: A simulation study. *Psychology, Crime & Law*, 8, 311–318. https://doi.org/10.1080/10683160208401822

Orthey, R., Vrij, A., Leal, S., & Blank, H. (2017). Strategy and misdirection in forced choice memory performance testing in deception detection. *Applied Cognitive Psychology*, 31(2), 139–145. https://doi.org/10.1002/acp.3310

Orthey, R., Vrij, A., Meijer, E., Leal, S., & Blank, H. (2018). Resistance to coaching in forced-choice testing. *Applied Cognitive Psychology*, 32, 693–700. https://doi.org/10.1002/acp.3443

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills*, *57*, 367–372. https://doi.org/10.2466/pms.1983.57.2.367

Plass, J. L., Moreno, R., & Brunken, R. (2010). *Cognitive load theory*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511844744

Shaw, D. J., Vrij, A., Mann, S., Leal, S., & Hillman, J. (2014). The guilty adjustment: Response trends on the symptom validity test. *Legal and Criminological Psychology*, *19*, 240–254. https://doi.org/10.1111/j.2044-8333.2012.02070.x

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118. https://doi.org/10.2307/1884852

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409. https://doi.org/10.1037/h0058700

Van Oorsouw, K., & Merckelbach, H. (2010). Detecting malingered memory problems in the civil and criminal arena. *Legal and Criminological Psychology*, *15*, 97–114. https://doi.org/10.1348/135532509X451304

Verschuere, B., Meijer, E., & Crombez, G. (2008). Symptom validity testing for the detection of simulated amnesia: Not robust to coaching. *Psychology, Crime & Law*, *14*(6), 523–528. https://doi.org/10.1080/10683160801955183