Malingering Detection in Memory Deficits: An Evaluation

Scott Alexander Chalmers

Malingering, and its study, is a topic in clinical neuropsychology that has generated a lot of controversy from testing the IQ of convicted murderers to litigation trails where plaintiffs demand compensation for mental or physical traumas incurred in their workplace. Here, three major tests: The Forced-Choice Test (Hiscock and Hiscock 1989), Test of Memory Malingering (TOMM) (Gansler et al.1995) and the Wechsler Memory Scales-Revised (WMS-R) test (Wechsler 1987) have been evaluated primarily on the basis of their capabilities to detect cases of genuine memory deficit as a result of head trauma from cases of malingering in clinical Neuropsychological settings. Tests such as these play a crucial role in a clinicians overall evaluation of a test subject. A poor test may result in a failure to detect cases of memory deficit malingering or detect malingering in genuine cases of memory deficits. The current review compares each test on the dimensions of sensitivity and specificity, the two vital dimensions of a test’s detection capability. It was found that conservative cut off points of test scores is a theme that persists throughout each test which means that malingering patients may successfully get away with faking their performance on a test. The implications of the conservative cuts offs are discussed.

INTRODUCTION

In Neuropsychological research, the distortion of test performance due to the presence of malingering by participants has become an important but controversial topic. In court cases, the final judgment may depend on the outcome of a clinical evaluation of an individual’s mental abilities and such an evaluation may include a test for malingering. Malingering, according to the American Psychiatric Association, is “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs” (American Psychiatric Association, 1994, p. 683) None the less, further refinement is needed to authenticate malingering diagnosis criteria for day to day use in the lab (Slick et al.1999). Malingering is not considered a psychological disorder but it does require the focus of clinicians when conducting studies where external incentives have been identified. In cases where tests of malingering are applied it is essential that these tests provide detective capability, which is the focus of the current review. The capability of three tests shall be examined throughout the duration of this discussion. The tests that will be evaluated include the ‘Forced-Choice Test’ (Hiscock and Hiscock 1989), the ‘Test of Memory Malingering (TOMM)’ (Gansler et al. 1995), a forced-choice test, and the ‘Wechsler Memory Scales-Revised (WMS-R)’ test, created by Wechsler in (1987). These methods will be evaluated based on the dimensions of sensitivity (a test’s effectiveness at detecting actual malingers) and specificity (a test’s effectiveness at ruling out non-malingering patients) in the detection of malingering. Other general implications that face psychologists when conducting research which involves malingering detection methodologies in the field, such as the practicalities of applying these methods within the testing environment shall also be highlighted. However, do note that these tests would be used in conjunction with other neuropsychological tests in a test battery, with thorough clinical interviews and reference to the past medical records of the test patient (Adams et al. 1996). Past research cited here for each test shall form the basis of this review and are discussed in each section. The purpose of this research is to evaluate whether the cut off points, where any scores registered above or below a certain cut off score is said to alert clinicians to possible malingering, are appropriate for determining the detective capability of each test. While no tests detective capability has 100% sensitivity or specificity it is hypothesised that the cut off points will be rather conservative meaning that specificity shall be higher compared to sensitivity on each test.

This investigation is constrained to looking at malingering detection tests in suspected head trauma patients, who are apparently suffering from memory deficits. The tests conducted are of a ‘known- group’ design, meaning the participants have been assigned groups, typically a health control group, with genuine head trauma patients and malingering simulators. A paper by Greve et al (2008) warns us that the correct allocation of malingers and non-malingers into separate groups is necessary, otherwise, the results, which may in future be used as typical results for these groups, will be skewed making the cut offs more conservative and not an accurate sample of that populations actual performance on a test (Greve et al. 2008). The cut off points of each test shall be evaluated on the basis of their malingering detection capability as deemed appropriate according to the literature.

1University of Aberdeen, King’s College, Regents Walk, Aberdeen AB24 3FX, United Kingdom

*To whom correspondence should be addressed:
scott.chalmers.09@aberdene.ac.uk
Forced-Choice Test
The Forced-Choice test procedure consists of trials asking the participant to perform simple memory tasks where they had to remember a 5 digit number. Although this test is simple it would appear to be very complex to anyone unfamiliar with the procedure. Meanwhile, they are asked to complete a record sheet containing multiple choice answers to questions based on the stimuli just presented. These tasks are then separated into a number of trials that the participant is told are increasing in difficulty. If the participant has genuine memory problems then we expect them to perform close to chance (which means getting about 50% of the questions right, which is the same as the odds of guessing the answers) and to score consistently across trials, if they are malingering then we expect them to perform below chance which means their scores get poorer as the trials get seemingly harder (Hiscock and Hiscock 1989).

The Forced-Choice test has proven to be very sensitive and specific at detecting cases of malingering, especially in comparison to an older test model, the Rey Fifteen-item test (1964), where 85% sensitivity and 97.5% specificity values were registered for an abbreviated version of the Forced-Choice test with a cut off point of 90% of the correct responses (Guilmette et al. 1994).

However, in a paper by Daniel Slick and colleagues, they reported that this cut off point is perhaps overly conservative as it only caught naive malingerers and more sophisticated malingerers could avoid detection, despite the specificity of the test being very high. When a less conservative cut off was used in their test (a shortened version of the Forced-Choice test called the Victoria Revision) a higher sensitivity of 80% was reported, but a lower specificity of 90% was also reported (Slick, Hopp, Strauss, Hunter, & Pinch, 1994). By using more conservative cut off points in the Forced-Choice test it appears that specificity increases, but sensitivity is sacrificed; on the other hand, when a more lenient cut off is used, the sensitivity can be increased but the specificity decreases. This trade-off is intrinsic to the Forced-Choice test because lowering the cut off point for the number of correct responses expected by patients means that less malingerers will be classified as not malingering (which is increased sensitivity), but in more severe cases of memory deficits that are atypical even of brain trauma patients who are not malingering are more likely to be falsely accused of doing so (which is decreased specificity). Slick et al (1994) also concluded that a shorter version with fewer trials (only 2) gave the same results and was more applicable in the lab. This means a shortened Forced-Choice test would be more practical than the full test constructed by Hiscock and Hiscock (1989).

Test of Memory Malingering
The ‘Test of Memory Malingering’ (TOMM) was designed by Tombaugh et al (1996) and is one of the most widely used tests for memory malingering in neuropsychology (Slick et al. 2004). In this test, 50 stimuli are presented for a total of 3 seconds per stimuli, after which three other trials follow: a recognition trial, a second recognition trial and then a delayed recognition trial. The last two trials are used to test for malingering (Merten et al. 2007).

The Test of Memory Malingering has notably poor sensitivity at determining cognitive ability (Tombaugh1996, cited Greve et al. 2006), but boasts high specificity (Merten et al. 2007). Greve et al (2008) showed that the cut offs for this test are too conservative. Though the specificity of the test is adequate, the sensitivity remains rather low due to conservative cut offs. However, when these cut offs are adjusted, greater sensitivity but lower specificity are reported. When compared to similar tests where sensitivity and specificity rates are about equal because of adjusted cut offs, the TOMM demonstrated the highest percentage rate of malingering detection compared to these other tests (Greve et al. 2008).

It is hardly surprising that the TOMM has such high specificity when even 5 year old children are able to perform as well as adults on the adult format of the test (Rienstra et al. 2010). None the less, we must call the validity of the test into question when its primary function is to detect malingering but it is not incredibly sensitive to cases of malingering. Gervais, Rohling, Green and Ford (2004) also concluded that the TOMM was insensitive to cognitive exaggeration, but they also concluded that anyone producing sufficient effort will not be incorrectly classified as malingering. This study revealed that when the TOMM is compared to other two-alternative forced-choice tests (the Computerised Assessment of Response Bias, CARB and the Word Memory Test, WMT), the TOMM proved to have better specificity, but was only useful at detecting extreme cases of malingering (Gervais et al. 2004). This is consistent with the idea that the TOMM is not very sensitive, but has high specificity. Still, the simple procedure does allow the test to be administered as part of a test battery with great ease.

Wechsler Memory Scales-Revised
The ‘Wechsler Memory Scales-Revised’ test was devised by Wechsler in 1987 and is not specifically a malingering detection test, but can be applied as such when cut off points for scores are introduced. The test relies on the tendency for malingering participants to perform poorly on the attention portion of the test and the memory portion of the test, whereas head trauma patients should perform better on the attention portion relative to the memory portion (Hom and Denney 2003). The difference score between the two is called the ‘General Malingering Index.’ The cut offs are then applied to this Index, a difference score that is sufficiently large enough can raise flags about a possible malingering participant (Iverson, Slick, and Franzen 2000).

Mitenberg et al (1993) found a sensitivity rate of 90% and a specificity of 77% on this test given the cut off scores used in that study (Mitenberg et al.1993 cited in Adams et al. 1996). The specificity of this study is very low particularly when contrasted with the Force-Choice Test and the TOMM. Iverson et al (2000) warns us of the possibility of errors in this test within the context of making clinical inferences. Because the rate of falsely detecting non-malingerers as malingerers is so high, any clinician should be dubious about its application. The study also recognised that once the cut offs where examined it was found
that using a more conservative cut off would increase sensitivity without significantly influencing specificity, but even here further examination of the patients flagged for malingering discovered that 5 out of 6 of them were suffering from other neurological disorders that affected their scores (Iverson et al. 2000). Clinicians must be cautious when interpreting the results of a single test.

CONCLUSIONS
Conclusively, under evaluation of these tests we can see that for a test to have high sensitivity, or high specificity, it must result in a reduction in the magnitude of the other. Whether due to errors by non-malingering patients or sophisticated malingering of those faking their results, there comes a point where the distinction between a legitimately poor scoring patient and a well versed malingerer becomes too hazy to determine, and consequently the nature of their answers based on a single test would falsely designate participants into each category on arbitrary terms. That is why these tests are delivered as part of a battery and in conjunction with clinical interviews and checks of past records as proposed by Adams et al (1996). The Forced-Choice Test proved to be highly sensitive and specific at detecting malingering but its cut off was too conservative. This conservative cut off point was a theme in the TOMM and the WMS-R as well. However, as a criticism of these tests, conservative cut off scores have a valuable use as they serve to avoid wrongly accusing non-malingers of malingering, which is a serious accusation. A general limitation of these tests is that, although they offer a method of detecting malingering suspects, they do not offer us an interpretive means of identifying malingers for sure. These results may alternatively be the result of insufficient effort and not malingering itself. Although, the TOMM and the Force-Choice Test rely on the malingerers achieving results that are statistically improbable given sufficient effort on the test has been provided. Tests specifically measuring effort are required just as well as malingering tests. Let us hope that the trade-off between sensitivity and specificity is only a lapse in the imagination of clinicians designing and executing these tests and not a necessary limitation of the methodology itself.

REFERENCES