Objective Tests of Symptom Exaggeration in Independent Medical Examinations

Jack Richman, MD
Paul Green, PhD
Roger Gervais, PhD
Lloyd Flaro, PhD
Thomas Merten, PhD
Robbi Brockhaus, PhD
David Ranks, PhD

Objective: *This study used the Medical Symptom Validity Test (MSVT) to examine exaggeration of memory impairment in disability claimants.* ***Methods:*** *The MSVT was administered to patients with soft tissue injuries undergoing an independent medical examination (IME). Their results were compared with those from groups of volunteers who were either trying their best on the test or simulating memory impairment.* ***Results:*** *Non-French-speaking volunteers, who were tested in French, showed near perfect performance on the effort subtests, but 42% of IME patients failed the effort tests in English. Their overall results were very similar to those of simulators.* ***Conclusion:*** *This study suggests that exaggeration of cognitive symptoms is widespread in disability-related evaluations. It would be unwise to accept self-reported memory complaints at face value. Criteria-normed symptom validity testing should be done to rule out symptom exaggeration.* (J Occup Environ Med. *2006;48:303–311*)

From AssessMed (Dr Richman), Toronto, Canada; Private Practice (Dr Green, Dr Flaro, Dr Gervais), Edmonton, Canada; Klinikum im Friedrichshain (Dr Merten), Berlin, Germany; Alexianer-Krankenhaus (Dr Brockhaus), Krefeld, Germany; and Private Practice (Dr Ranks), Salt Lake City, Utah.

Address correspondence to: Jack Richman, MD, 5925 Airport Road, Suite 400, Mississauga, Ontario, Canada L6J 4C1; E-mail: jrichman@assessmed.com.

DOI: 10.1097/01.jom.0000183482.41957.c3

If a patient engages in symptom exaggeration, it is usually difficult to identify the exaggeration and, until recently, it was impossible to quantify it objectively. Wolfe[1] reported that six rheumatologists classified a woman as being disabled by fibromyalgia. However, after viewing a videotape showing her leading an active life, they all agreed that they had been mistaken. A detailed review of records by others revealed "a pattern of lies and deception on the part of the plaintiff." Wolfe[1] pointed out that physicians have a "bias toward trust and empathy" but, in the latter case, "Everyone got it wrong." It was concluded that the reliability of ratings of disability by rheumatologists is poor, and it was implied that medical specialists cannot reliably identify symptom exaggeration.

Physicians assessing claimed disability are typically forced to use judgment to evaluate self-reports of subjective symptoms such as memory problems, headaches, back pain, sleeplessness, loss of sex drive, decreased interests, fatigue, sadness, mental confusion, and many other symptoms. Although some of these symptoms could be disabling, their severity cannot be verified objectively, independently of the patient's self-reports or the reports of others. Symptom exaggeration can go undetected, especially if there are incentives to appear impaired such as in personal injury lawsuits, disability insurance claims, or in the evaluation of soldiers claiming disability to avoid posting to a war zone. This can lead to incorrect conclusions about disability, as in the case cited by Wolfe.[1]

Memory complaints are typically present in people claiming disability from fibromyalgia, major depression, soft tissue injury such as whiplash, chronic fatigue, chronic pain, anxiety, and many other conditions. Memory complaints were actually more prevalent in such patients than in cases of established brain disease on a standardized memory complaints inventory.[2,3] Self-reports of memory complaints often overestimate the person's actual memory impairment, especially when there is some external incentive for symptom exaggeration. However, it has been shown that exaggeration of memory problems can be measured objectively using tests known as symptom validity tests (SVT) or effort tests. If such effort tests are failed, formal test results are of questionable validity and subjective self-reports are doubtful. Green et al[4] reported that memory test scores were suppressed to a much greater degree by poor effort than they were by severe traumatic brain injuries, strokes, or tumors in a large outpatient sample. Similar results were obtained independently by Constantinou et al,[5] who found that the effect of effort on neuropsychologic test scores in cases of mild head injury was so large that effort explained 47% of the variance in the test battery. Because effort has such a profound impact on ability test scores and because symptom exaggeration often occurs in the context of compensation or disability claims, effort tests are now required in any formal memory evaluation, especially when there are secondary gains for being impaired.[6]

Exaggeration of memory problems may be measured objectively using the Word Memory Test (WMT),[7] a verbal memory test with built-in effort subtests. Gervais et al[8] administered the WMT to patients with fibromyalgia attending a rheumatology clinic. The patients with fibromyalgia were divided into three groups: 1) those with no disability claim, 2) those with an existing disability pension, and 3) those currently applying for a disability pension. The failure rate on the effort subtests of the WMT was 11 times higher in those with an unsettled disability claim (44%) than in those with no such claim (4%). The effort test failure rate in those with an existing disability claim was intermediate between those of the latter two groups (23%). Memory complaints were widespread in all groups but, in those who passed effort tests, actual memory scores were within the normal range. In those who failed effort testing, the mean memory test scores were very impaired but of doubtful validity. Neither the diagnosis nor the severity of disease could explain the differences in effort test failure rates between groups. It was concluded that the presence of a disability claim was the major factor explaining the differences in effort test failure rates.

Researchers in an epilepsy brain surgery clinic[9,10] compared WMT scores in a group of patients with psychogenic nonepileptic seizures (PNES) and patients with epilepsy who were about to undergo brain surgery. More than half of the PNES cases failed effort testing, whereas very few of the cases with true seizure disorders failed. If a patient failed the effort subtests, they were 20 times more likely to belong to the PNES group than to the epilepsy surgery group. Those PNES cases who failed effort testing were 63 times more likely than the patients with confirmed epilepsy to display significant cognitive impairment on a wide range of cognitive tests. Thus, effort test failure was found most often in those with the least demonstrable brain pathology, and it was linked with invalid cognitive test results.

The Medical Symptom Validity Test[11] is a Microsoft Windows-based computerized test, which is similar to the WMT and was designed for use by physicians. It is a very brief automated memory screening test with integral effort measures. The MSVT requires approximately 5 minutes of patient time on task, which is much less time than needed to administer the WMT. The test is almost entirely self-administered on a computer and can easily be given by a nurse. Poor effort is detected automatically and, if effort is poor, the computer produces a report, warning that the two memory subtest scores are probably not valid. The examinees are asked at the end of the test whether they made a full effort, and their "yes" or "no" responses are entered into the computer. If their effort scores are low but they claim to have made a full effort, there is a contradiction, from which we must conclude either 1) that there is evidence of extremely severe impairment, consistent with moderate to severe dementia and probable inability to live independently in safety; or 2) when there is no dementia, as in most cases, there is exaggeration of memory problems. If effort is good, the memory test scores could be either valid and in the normal range or valid and lower than normal. In the latter case, depending on the clinical context and history, further investigation might be considered.

In a study of the accuracy of the MSVT in Germany, using a German adaptation of the test, Merten et al[12] found that the MSVT classification of good versus poor effort was very similar to that derived from another well-established effort test, the Amsterdam Short Term Memory Test.[13] The MSVT subtests were 97% to 100% accurate in differentiating between good and poor effort in this simulator study.

The purpose of the current pilot study was to evaluate effort, and, by inference, symptom exaggeration in a series of 106 cases undergoing an independent medical examination (IME). They were mostly cases of soft tissue injury and were being assessed by physicians of various specialties in the context of an insurance disability claim or personal injury claim. The primary question was how many people, who were seen for an IME by an MD and were complaining of cognitive problems, often secondary to pain, would be found to show poor effort on the MSVT, which was administered by a nurse? For comparison, data were gathered from English-speaking volunteer adults and children using the same test but in French. Some of the volunteers were fluent in French, whereas others did not speak any French. Data from various other groups reported in the MSVT test manual[11] were used to aid interpretation of the current results.

Materials and Methods

Participants

Disability Claimants. The mean age of the claimants was 42 years (standard deviation [SD], 12) and 36% were men. All were referred to AssessMed (a private medical organization specializing in the performance of assessments of disability, including IMEs) for an IME for evaluation of disability for a third party such as an insurance company, workers' compensation board, or a lawyer. The majority of claimants (68) were classified as cases of soft tissue injury or fibromyalgia. There were 20 cases with orthopedic injuries, other than soft tissue injuries, 10 cases of chronic pain, two cases of severe traumatic brain injury, two with mild head injuries and one with head injury of unknown severity. There was one case of rheumatoid arthritis, one chronic fatigue syndrome, and one with a primary diagnosis of major depression.

This was a practical pilot study in a series of four clinics in which physicians performed IMEs. The intention was that a consecutive series of cases would be tested. However, practical matters intervened such as certain clinics having the computer equipment before others, nurses being absent, one client having less than a grade three reading level, and some physicians learning and implementing the method before others. Also, people were not tested if they came to the appointment with an interpreter ($n = 49$) because it was felt that language could be a confounding variable. As a result, the patients tested, although not technically a consecutive series, were consecutive if the following conditions were met: 1) that there were cognitive complaints as a result of their symptoms, 2) that they spoke and read English, and 3) that nurse training was complete and software was installed. One hundred six cases out of a total of 296 undergoing IMEs in the AssessMed clinics met these criteria between September 2004 and January 2005, which provides some preliminary information on MSVT failures in IME candidates.

Child and Adult Volunteers. Normative data for the French MSVT were obtained by one of the authors (RG), who tested a sample of 45 community volunteers under standardized conditions. The sample was 56% female and ranged from 7 to 58 years of age (mean, 25.3 years; SD, 15.3). There were 19 children aged 7 to 16 years (mean, 10.3 years; SD, 2.5) and 26 adults (mean, 37.7; SD, 9.8). Eleven (58%) of the children and 12 (46%) of the adults spoke French as a first language or had sufficient familiarity with the language to read and understand the on-screen instructions. The other participants were non-French-speaking. The children in the sample were at their age-appropriate grades in school. All the adults had completed at least high school-level education, with 15 (58%) having a bachelor's degree and five (19%) holding a master's degree.

The participants were tested under three conditions using the MSVT in French:

1. Thirteen French-speakers were asked to try their best on the MSVT, in which all stimuli on the computer were in French.
2. Twenty-one non-French-speakers were also asked to try their best on the MSVT, in which all stimuli were in French.
3. Finally, 13 French-speaking adults and three previously tested older French-speaking children were asked to fake memory impairment on the French MSVT, in a manner that would be credible to the examiner, without being detected as exaggerating. To assist them in assuming this response set, we asked participants to imagine that they had been involved in a motor vehicle accident in which they had sustained a flexion–extension neck injury or "whiplash." They were to imagine that they were now claiming to have memory impairment that was compromising their ability to work and they were pursuing litigation for compensation. If testing confirmed the alleged memory complaints, they should imagine the possibility of a substantial financial settlement.

Groups Tested in Previous Validation Studies. Data from several independent researchers were included in the normative comparison groups, which are reported in the MSVT program and test manual[11] and reproduced in Tables 1 through 3. These studies showed that passing the MSVT effort measures requires minimal ability and that most people making an effort will easily pass the effort subtests. Normal adult volunteers were tested in English by Ranks, in French by Gervais, and in German by Merten (Table 1). Volunteer simulators were tested in French by Gervais (Table 2) and in German by Merten (Table 3). Adult patients of mixed diagnoses were tested clinically in English by Green and Gervais (Tables 1 and 3). Children were tested clinically in English by Flaro (Tables 1 and 3) and experimentally in French by Gervais (Table 2). Patients with dementia were tested in German by Brockhaus (Table 3).

Test Procedure

The Windows-based MSVT was used to test all participants. A nurse or psychologist read aloud the instructions, which ask the person to watch the screen while a list of word

TABLE 1
Mean Medical Symptom Validity Test (MSVT) Immediate Recognition and Delayed Recognition Scores Are Between 96% and 99% Correct in All Groups Tested in Their First Language and Assumed to be Making a Full Effort

Group	No.	IR	DR	CNS	PA	FR
Adult outpatients passing WMT (English oral MSVT, Green)	92	99% (3)	99% (3)	98% (5)	94% (11)	73% (15)
Adult outpatients passing WMT (English computer MSVT, Green)	47	98% (3)	97% (5)	95% (7)	94% (10)	69% (15)
Adult outpatients passing WMT (English computer MSVT, Gervais)	81	99% (3)	99% (3)	98% (4)	93% (10)	74% (15)
Adults with severe TBI or neurologic disease passing WMT (English oral MSVT, Green)	26	98% (3)	98% (4)	96% (6)	89% (15)	69% (16)
Adults with severe TBI or neurologic disease passing WMT (English computer MSVT, Green)	14	98% (3)	96% (5)	95% (6)	94% (10)	61% (15)
Children tested clinically passing computer WMT (English oral MSVT, Flaro)	28	100% (2)	99% (3)	98% (3)	96% (8)	67% (22)
German university students with good effort (German oral MSVT, Merten)	18	99% (4)	99% (2)	99% (3)	96% (6)	91% (8)
Children tested clinically, mean age 12, mean FSIQ 88, passing WMT (English oral or computer MSVT, Flaro)	50	99% (3)	98% (4)	97% (4)	96% (7)	70% (22)
Adult volunteers with mean 17-yr education (English MSVT, Ranks)	10	99% (2)	99% (2)	99% (2)	99% (3)	72% (10)

IR and DR indicate immediate and delayed recognition, which is a forced choice task; CNS, consistency of responses from IR to DR; PA, paired associate recall; FR, free recall of words; WMT, Word Memory Test; TBI, traumatic brain injury; FSIQ, Full-Scale Intelligence Quotient.

TABLE 2
Mean Medical Symptom Validity Test (MSVT) Immediate Recognition and Delayed Recognition Scores Were No Different in Children or Adults Tested in a Foreign Language (French) versus Children Tested in Their Own Language (French)

Group	No.	IR	DR	CNS	PA	FR
Fluent-French-speaking child volunteers given the French computer MSVT (Gervais)	12	98% (5)	99% (2)	98% (5)	84% (31)	74% (25)
Non-French-speaking child volunteers given the French computer MSVT (Gervais)	8	97% (6)	98% (4)	95% (8)	27% (23)	42% (13)
Non-French-speaking adult volunteers given the French computer MSVT (Gervais)	8	98% (4)	99% (2)	98% (4)	46% (25)	54% (14)
Significance of differences between groups		NS	NS	NS	0.001	0.01

IR and DR indicate immediate and delayed recognition, which is a forced choice task; CNS, consistency of responses from IR to DR; PA, paired associate recall; FR, free recall of words.

pairs is presented twice at a rate of 6 seconds per pair. Then, the computer successively presents one word from each pair, which was shown previously and one which was not shown, and the person is required to select the word shown previously in the original list. This produces a total of 20 test items on the immediate recognition trial (IR). After completion of this subtest, 10 minutes is occupied with activities not linked with verbal memory. Then the same recognition testing is performed again using different foil words in the delayed recognition trial (DR). One would expect that people obtaining either a correct answer or an incorrect answer the first time (IR) would be consistent on the second trial (DR). However, the claimants who failed the IR or DR trials would often get an incorrect answer the first time but get it correct the second time or vice versa. A consistency score (CNS) between the two trials is calculated by the computer.

After the DR trial, testing proceeds to the paired associates trial (PA), in which the person is asked to sit where the computer screen cannot be seen. The tester reads the first word of each pair from the screen and the person is asked to say the word that went with it in the original list. Finally, the person is asked to recall as many words as possible from the original list in the free recall trial (FR). The tester records the person's responses using the computer. Computer scoring of the person's responses is automatic and reporting is available in several graphic and numeric formats contrasting the single case scores with those from numerous comparison groups.

Results

Good Effort Groups

It is very clear in Table 1 that, in all groups, the mean scores were very close to 100% correct on both of the MSVT recognition measures, IR and DR, which were designed to measure effort. The median score on these subtests for all groups in Table 1 was 100% correct irrespective of whether the groups consisted of adults or children. The range of mean recognition scores in Table 1 was 95% to 100% correct. There was no difference in mean MSVT recogni-

TABLE 3
Mean Medical Symptom Validity Test (MSVT) Scores in Independent Medical Examination (IME) Patients, Compared With Adults Who Were Asked to Fake Memory Impairment, Adult Compensation Claimants Failing the MSVT Effort Subtests, Cases of Advanced Dementia and Children With Mentally Retarded Intelligence

Group	No.	IR	DR	CNS	PA	FR	MSVT Test Type and Examiner
IME cases passing MSVT effort subtests	61	99% (2)	99% (3)	98% (4)	94% (8)	72% (18)	English/computer MSVT, Richman
IME cases failing MSVT effort subtests	45	72% (18)	70% (18)	67% (15)	60% (19)	42% (18)	English/computer MSVT, Richman
All IME cases	106	87% (18)	87% (18)	85% (18)	80% (22)	59% (23)	English/computer MSVT, Richman
Children with mean full-scale intelligence of 65	7	95% (8)	95% (6)	95% (4)	87% (13)	67% (24)	English/computer MSVT, Flaro
Patients with advanced dementia	14	72% (19)	72% (18)	74% (15)	33% (23)	11% (12)	German/oral MSVT, Brockhaus
Patients with early dementia	48	88% (13)	89% (13)	84% (13)	57% (26)	33% (22)	German/oral MSVT, Brockhaus
Adult outpatients with poor effort (fail WMT)	32	77% (22)	75% (19)	75% (18)	62% (21)	42% (17)	English computer MSVT, Green
Volunteers faking memory impairment	18	65% (12)	61% (16)	62% (15)	45% (13)	38% (14)	German oral MSVT, Merten
Volunteers faking memory impairment	11	65% (32)	60% (21)	65% (15)	40% (19)	38% (16)	French computer MSVT, Gervais

Note: Those who were asked to fake memory impairment scored lower than patients with dementia on IR and DR effort subtests but much higher than patients with dementia on PA and FR, the memory subtests.

IR and DR indicate immediate and delayed recognition, which is a forced choice task; CNS, consistency of responses from IR to DR; PA, paired associate recall; FR, free recall of words; WMT, Word Memory Test.

tion scores between university students trying their best and children who had various clinical conditions and a mean IQ of 88. There was no difference in mean MSVT recognition scores between adults with severe traumatic brain injury and healthy adults with high educational levels.

Even when children were tested on the MSVT in a language that they did not understand but based on the Latin alphabet (French), their median score was 100% correct on the IR and DR subtests and their mean scores were close to 100% correct. In Table 2, it is shown that the mean IR and DR subtest scores were 97% and 98% correct in non-French-speaking children who were tested in French. These were not significantly different from the mean scores of 98% and 99% correct in children whose first language was French. Adult non-French-speakers also obtained mean scores of 98% and 99% correct on MSVT IR and DR subtests, scores that are no different from those of the French-speaking children. Thus, people can perform almost perfectly in recognizing words on IR and DR (the effort subtests) even if they do not understand the meanings of the words and are unable to pronounce them correctly.

In contrast, as expected, the non-French-speakers performed significantly worse than the French-speaking children when required to perform the true memory subtests (paired associates and free recall, $P < 0.001$), which do involve understanding word meanings. In PA, the person was given the first word of the pair (eg, "chocolat" pronounced by a fluent French-speaker) and was asked to say the second word that was paired with it (eg, "chaud"). Credit was given if the word was not pronounced correctly but resembled the correct word phonetically (eg, "chode" for "chaud" and "chocolate" instead of "chocolat"). Similarly, FR of words on the list in which understanding of meaning assists with memory was far superior in the French-speakers than in the non-French-speakers ($P < 0.001$), although the French-speakers were younger (French mean age, 15.5 years [SD, 11.6]; non-French mean age, 26 years [SD, 16]; $P < 0.05$). Thus, in summation, for the effort/recognition portion of the MSVT (IR and DR), knowing the meanings of the French words and having prior exposure to them was not necessary for children to score 100% correct. However, as we would expect, a lack of knowledge of the language was a handicap in performing the paired associates and free recall subtests, which are sensitive to differences in memory ability.

Another sign of the relative insensitivity of the MSVT recognition subtests to actual impairment is the fact that children with a full-scale intelligence score less than or equal to 70 (mean, 65; SD, 5) scored a mean of 95% correct on the MSVT IR and DR effort subtests (Table 3). Of those groups assumed to be trying their best, only patients with advanced dementia obtained a mean score below the 85% cutoff on MSVT IR and DR subtests. In Table

3, cases of advanced dementia in permanent institutional care in Germany, with a mean age of 78 years, obtained mean scores of 72% correct on MSVT IR and DR subtests, respectively, when tested with the German adaptation of the MSVT.

Poor Effort and Simulators

In several of the groups shown in Table 1, outpatients were selected because they were assumed to be making a full effort as shown by passing the effort subtests of a different test, the WMT.[7] In contrast, outpatient groups from the same series, selected for failing WMT effort subtests, are shown in Table 3. Their MSVT scores were markedly lower than those of the patients who had passed WMT effort testing (Table 1).

The French simulator group scored significantly lower on all MSVT subtests than the French good effort groups ($P < 0.0001$), and all members of the French simulator group failed the MSVT effort subtests (IR and DR). They produced scores very similar to the German simulators on all subtests (Table 3). There was almost no overlap between the effort scores of the two simulator groups versus those of the good effort volunteers in this study whether tested in English, French, or German. All simulators failed the MSVT effort subtests and were detected as showing incomplete effort. With only one exception, all good effort volunteers tested in French or German scored greater than 85% correct on IR, DR, and consistency. In Table 3, it is clear that, in all instances, the mean scores of the two simulator groups were lower than those of the advanced dementia group on the easiest subtests (IR and DR). Conversely, on the harder subtests (PA and FR), the mean scores of the two simulator groups were all higher than those of the advanced dementia group. This pattern allows us to distinguish the simulators from the advanced dementias.

Medical Symptom Validity Test Results in Claimants Undergoing an Independent Medical Examination

Sixty-one IME cases passed the MSVT effort subtests and 45 failed. That is, in 45 cases, the mean MSVT IR, DR, or consistency scores were at or below 85% correct. The 61 IME cases passing MSVT subtests scored like most other groups who were assumed to be making a good effort. They obtained mean scores of 99% correct on IR and DR subtests (Table 3). The mean of IR, DR, and consistency was calculated for each case. Of those who passed the MSVT, 43 cases scored a mean of 100%, 10 scored a mean of 97%, eight scored a mean of 93% and none scored 90% or lower, whereas the highest mean score in those failing the MSVT was 90%. Hence, there was no overlap between the mean effort scores of those who passed the MSVT and those who failed the MSVT. Thus, it is clear that the 85% cut off for failure is a realistic value to detect simulation in all cases in which dementia is not a factor.

None of the claimants undergoing an IME had been previously diagnosed with dementia and they did not present with a clinical picture of dementia, yet 45 cases failed the MSVT. In the IME cases who failed the MSVT, 1) the mean scores on the effort subtests (mean IR = 72% and DR = 70%) were very similar to the mean scores from people with advanced dementia tested in an institution (mean IR = 72% and DR = 72%); but 2) the mean scores on the harder subtests, PA and FR, were significantly and paradoxically higher than in the dementia group (Table 3; the latter differences are significant at $P < 0.01$ or lower on Mann-Whitney U tests). On IR, the IME cases failing the MSVT scored 13.5 standard deviations lower than the mean from those who passed the MSVT. On DR, those who failed the MSVT scored more than 9 standard deviations lower than the mean from those who passed the MSVT. Although these data are not normally distributed, such figures illustrate the extreme difference between those who passed and those who failed the MSVT.

In those failing the MSVT, the mean MSVT IR score was extremely low. For example, it was more than two standard deviations lower than that of children with a mean IQ of 65 (Table 3). It was more than four standard deviations lower than the mean scores from English-speaking children tested in French and who did not understand French (Table 2). The IME cases failing the MSVT produced scores most similar to those of people who were asked to simulate or fake memory impairment in two experimental studies (Table 3). There was no overlap between the effort scores of IME cases who failed the MSVT and the scores from children tested in a foreign language with the MSVT. None of the latter cases scored as low as 85% on IR, DR, or consistency. Hence, language was not a barrier to a near perfect performance on the effort subtests in this study, in which the participants knew the letters of the Latin alphabet but could not understand the words.

The two most numerous diagnostic groups in this sample were 68 cases of soft tissue injury or fibromyalgia, 33% of whom failed the MSVT, and 20 cases with orthopedic injuries other than soft tissue injuries, of whom 50% failed the MSVT. In the IME sample, using a cutoff of 85% or below to define poor effort, there was agreement in 88% of cases between the MSVT IR and DR scores with regard to good or poor effort. There was 94% agreement between classifications of good versus poor effort using the DR subtest and the score for consistency of responses between IR and DR subtests. Despite restricted variance, the correlations were as follows: IR and DR, $r = 0.8$; DR and consistency, $r = 0.8$; IR and consistency, $r = 0.87$. There was no difference between the mean years of education in adults who passed the MSVT (mean, 13.2; SD, 2) versus those who failed the MSVT (mean, 12.7; SD, 2).

Discussion

In most people, the MSVT recognition subtests provide objective measures of effort as opposed to ability. English-speaking children, who spoke no French, took the test in French and they scored 100% on the recognition memory subtests in most cases. Children with intelligence in the mentally retarded range scored a mean of 95% correct (Table 3). Patients with severe traumatic brain injuries or neurologic diseases scored a mean of 98% correct. The only people found to have genuine difficulty with the MSVT recognition subtests were patients with dementia sufficient to render them incapable of living independently.

At the end of the test, every person undergoing an IME in the current study was asked whether they had made a full effort on the test. In all cases, their response was entered as "yes." Nevertheless, 42% of cases failed the test. The failures scored much lower than mentally handicapped children on the recognition measures. They scored so much lower than children tested in a foreign language and compared with brain-injured patients that their scores cannot be accepted as valid. Most of those who passed the MSVT scored either 100% or 97% correct on the mean of the effort measures. Yet, on average, the IME cases failing the MSVT obtained scores on the recognition subtests, as low as those from patients with advanced dementia (Table 3). Therefore, one must doubt their report of making a full or even reasonable effort.

On the MSVT recognition subtests, groups of people who were asked to simulate memory impairment also scored as low as or lower than people with dementia (Table 3). Signs of inconsistent effort emerge when we examine the scores of simulators on the very easy MSVT subtests, the recognition memory subtests (IR and DR), compared with the more difficult subtests (PA and FR). The groups simulating impairment scored lower than the patients with dementia on the easiest subtests (IR and DR), but they scored higher than the patients with dementia on the more difficult subtests (PA and FR). This is extremely unlikely in someone making a consistent effort and it is contrary to the objective difficulty of these tasks. In patients with advanced dementia, the FR subtest score, a reflection of memory, was a mean of only 11% correct, less than one sixth of their mean IR score of 72% correct (a reflection of effort). These scores reflect the actual relative difficulty levels of these subtests for people of very low abilities. Yet, in the German or French simulators in Table 3, the FR scores (38% correct) were approximately two thirds as large as their mean IR scores (65%). If the simulators were truly unable to score as highly as patients with dementia on the very easiest recognition memory subtests, they would not be expected to score significantly higher than the patients with dementia on the most difficult subtest, FR. Yet they did so. If the results of the IR and DR were a true reflection of their abilities (ie, as poor as or worse than the patients with dementia), they could not be considered mentally competent to advise a lawyer and would probably require institutionalization.

The confusing pattern of test scores in simulators is easily explainable. We know that the simulators' test scores do not reflect their actual abilities because we asked them not to try their best for experimental purposes. We know that these volunteers were making a deliberate effort to produce impaired MSVT scores but to avoid being identified as doing so. Therefore, their test scores do not reflect their actual abilities or the objective difficulty levels of the various MSVT subtests. It is most likely that the same explanation applies to the IME cases who failed the MSVT. Just like the simulators, the latter cases produced scores as low as patients with advanced dementia on the easiest subtests while scoring higher than patients with advanced dementia on the most difficult subtests. Like the known simulators, their scores do not reflect the objective task difficulty of the various MSVT subtests.

Considering the intrinsic inconsistencies in their scores across MSVT subtests and contrasting their results with those of other groups, it is hard to think of any explanation of the lowered MSVT test scores in the IME cases failing the MSVT apart from deliberately poor effort. Yet all these cases stated that they had made a full effort on the MSVT. We know with a high level of confidence that these people were not making a full effort on the MSVT but they claimed to be doing so in the context of an IME. This raises a question about the accuracy and reliability of their self-reporting in general and especially the reliability of their symptom reports in the IME.

It might be argued that failure on the MSVT effort subtests reflects some intrinsic variable such as depression, pain, or actual cognitive impairment. However, such arguments must be rejected because they do not explain the group data. The MSVT subtests were designed to be even easier than the equivalent subtests of the WMT, which has twice as many word pairs, and it has already been shown that severe brain injury cannot explain failure on the WMT. In fact, WMT effort test failure rates are greater in those with the least degree of brain injury than in those with the most severe brain injuries.[14] Those with the most objective damage to the brain on computed tomography or magnetic resonance imaging scans actually fail the WMT effort subtests much less often than those with normal brain scans.[15] Depression and chronic pain cannot explain WMT effort test failures,[16,17] and so they cannot explain failure on the even easier effort subtests of the MSVT. Also, the symptoms of fibromyalgia cannot explain failure on the MSVT. It was shown that the main factor predicting failure on the WMT effort subtests in groups with fibromyalgia was not reported pain or disability, but the presence of a financial claim for disability.[8]

If we were to accept that pain, depression, or symptoms of fibromyalgia actually caused failure on the MSVT effort subtests, we would be in the difficult position of having to explain how such factors could cause scores in the current study so much lower than those of mentally retarded children, English-speaking children tested in a foreign language, and brain-injured patients. We would have to explain why the pattern of test scores in MSVT failures was so similar to that of people who were asked to simulate (fake) memory impairment in experimental studies. That is, we would have to account for recognition memory scores as low as those of patients with dementia but free recall scores higher than those of patients with dementia. The best explanation is that those IME cases failing the MSVT were almost certainly making a poor effort, although they all denied making a poor effort in the context of a disability claim. It seems most likely that they were inclined to exaggerate their cognitive difficulties to support their claims.

This study shows that cognitive symptom exaggeration can be measured objectively in an IME setting and that it occurs in a substantial subgroup of claimants with soft tissue injuries. If a person chooses to fail an easy memory test, which we know they can pass, and then states that they tried their best, we must conclude that the person's self-reported memory complaints cannot be relied on. It has already been shown that poor effort of this type is associated with a very major and widespread suppression of test scores on other cognitive tests, including but not restricted to memory tests.[4,5] Further studies are needed to examine other self-reported symptoms in those who pass versus those who fail effort tests in an IME setting. We would hypothesize that those who fail such effort testing are probably exaggerating their other symptoms.

Symptom exaggeration can create a seriously misleading impression of impairment and disability, which is relevant at all stages of medical evaluation and treatment. It would be desirable to study groups of patients prospectively before they enter into rehabilitation programs or other treatments designed to minimize their disability and restore the ability to work. It is quite likely that effort test failure will represent a negative prognostic factor with regard to a return to work.

The patients in the current study endorsed memory problems before being tested and they were encouraged by the nurses and doctors to try their best on the MSVT. Although claiming to have tried their best, many patients did not do so, indicating a tendency not to comply with advice in a medical setting. It is a logical step to anticipate that such a lack of compliance will also be manifested in failure to follow treatment advice and a tendency to bias self-reports of symptoms when reporting on their progress over time to both their physicians and their insurance companies or agencies such as the workers' compensation board.

Undetected symptom exaggeration is not just of economic interest to insurance companies, although that is a consideration with major financial implications. It could also be a major contaminating variable in clinical studies of the relative effectiveness of alternative forms of treatment to the extent that self-reported function or symptom reporting is assessed. Exaggerated symptoms could lead to unnecessary treatments and possibly, in some cases, to severe adverse side effects, especially when there are no objective medical findings and the diagnosis relies on self-reported symptoms, as in the case of chronic pain, for example. Acknowledging the high risk of symptom exaggeration in such cases, Waddell[18] has recently argued that all people claiming compensation for chronic pain should be given some form of "effort testing" and that consideration must be given to "motivation" during the evaluation or a lack thereof.

The current study is the first step toward introducing well-validated, objective tests of effort into medical disability evaluations but it has many limitations. One limitation to this study is that there was some sampling bias, which could not be avoided. In retrospect, it is unknown how many from the whole sample would have passed the MSVT effort subtests, if they had all been tested. However, in the unlikely event that all those not tested would have passed the effort tests, the observation of 45 failures means that, at the very least, 15% of the whole IME sample would be exhibiting poor effort on the MSVT and, by inference, symptom exaggeration. Also, this study did not incorporate any independent variables such as treatment outcome measures, compliance with treatment, or other correlates of MSVT effort test failure. Further studies are needed to determine the prevalence of symptom exaggeration in samples of cases undergoing IMEs and to study the correlates of effort test failure in those undergoing treatment.

Conclusion

This study suggests that exaggeration of cognitive symptoms is widespread in disability-related evaluations. It would be unwise to accept self-reported memory complaints at face value. Criteria-normed symptom validity testing should be done to rule out symptom exaggeration.

References

1. Wolfe F. For example is not evidence: fibromyalgia and the law [Editorial]. *J Rheumatol.* 2000;27:1115–1116.
2. Green P. *Memory Complaints Inventory for Windows* (computer program). Edm-

onton, Canada: Green's Publishing; 2005;11:139–145.

3. Green P, Gervais R, Merten T. Das Memory Complaints Inventory (MCI): Gedächtnisstörungen, Beschwerdenschilderung and Leistungsmotivation [The Memory Complaints Inventory (MCI): memory impairment, symptom presentation, and test effort.] *Neurologie & Rehabilitation.* 11:132–138.
4. Green P, Rohling ML, Lees-Haley PR, Allen LM. Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury.* 2001;15:1045–1060.
5. Constantinou M, Bauer L, Ashendorf L, Fisher J, McCaffrey RJ. Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tasks? *Arch Clin Neuropsychol.* 2005;20:191–198.
6. Iverson GL, Binder LM. Detecting exaggeration and malingering in neuropsychological assessment. *J Head Trauma Rehabil*. 2000;15:829–858.
7. Green P. *Green's Word Memory Test for Windows: User's Manual*. Edmonton, Canada: Green's Publishing; 2003.
8. Gervais RO, Russell AS, Green P, Allen LM, Ferrari R, Pieschl S. Effort testing in patients with fibromyalgia and disability incentives. *J Rheumatol*. 2001;28:1892–1899.
9. Williamson DJ, Drane DL, Stroup ES, Miller JW, Holmes MD, Wilensky AJ. Detecting cognitive differences between patients with epilepsy and patients with psychogenic nonepileptic seizures: effort matters. *Epilepsia.* 2004;45(suppl 7):179.
10. Williamson DJ, Drane DL, Stroup ES, Miller JW, Holmes MD. Most patients with psychogenic nonepileptic seizures do not exert valid effort on neurocognitive testing. *Epilepsia.* 2003;44(suppl 9):13.
11. Green P. *Manual for the Medical Symptom Validity Test for Windows.* Edmonton, Canada: Green's Publishing; 2004.
12. Merten T, Green P, Henry M, Blaskewitz N, Brockhaus R. Analog validation of German language symptom validity tests and the influence of coaching. *Arch Clin Neuropsychol*. 2005;20:719–726.
13. Schmand B, Lindeboom J. *Amsterdam Short-Term Memory Test Manual.* Leiden, The Netherlands: PITS; 2005.
14. Green P, Iverson GL, Allen LM. Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*. 1999;13:813–819.
15. Green P, Lees-Haley PR, Allen LM. The Word Memory Test and the validity of neuropsychological test scores. *J Forens Neuropsychol*. 2002;2:97–124.
16. Rohling ML, Green P, Allen LM, Iverson GL. Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Arch Clin Neuropsychol*. 2002;17:205–222.
17. Gervais R, Green P, Allen LM, Iverson GL. Effects of coaching on symptom validity testing in chronic pain patients presenting for disability assessments. *J Forens Neuropsychol.* 2001;2:1–19.
18. Waddell G. *Compensation for Chronic Pain*. London: The Stationary Office; 2004.