Child Neuropsychology
2003, Vol. 9, No. 3, pp. 189–207

0929-7049/03/0903-189$16.00


# Word Memory Test Performance in Children

Paul Green[1] and Lloyd Flaro[2,*]
[1]Neurobehavioural Associates, Edmonton, Alta., Canada
and [2]Private Practice, Edmonton, Alta., Canada

## ABSTRACT

One-hundred and thirty-five children between the ages of 7 and 18 years were evaluated clinically. Their diagnoses included Fetal Alcohol Syndrome (FAS) or Effects, Schizophrenia, Bipolar Mood Disorder, various neurological diseases, Attention Deficit Hyperactivity Disorder (ADHD), Conduct Disorder, Oppositional-Defiant Disorder and learning disabilities. As part of a comprehensive neuropsychological assessment, the children were given the Word Memory Test (WMT; Green, Allen, & Astner, 1996; Green & Astner, 1995), containing various subtests which measure, respectively, effort and verbal memory. Although age and verbal intelligence are known to affect scores on most ability tests, they were not found to be significant determinants of WMT effort scores. Younger children did not score any lower on the effort subtests than older children. The children scored as well as a group of parents seeking custody of their children and they scored higher than adult patients with mild head injuries. The computerized WMT requires some basic reading skills and some children with lower than a grade 3 reading level scored at a relatively low level on the effort subtests. The current data suggest that most children with at least a grade 3 reading level can pass the WMT using the adult criteria. It is concluded that the WMT is potentially useful in the evaluation of effort during pediatric neuropsychological evaluations. Further research is needed to replicate these findings and to develop child norms for the memory subtests.

## INTRODUCTION

There is now a large literature examining the way in which the amount of effort applied to testing influences neuropsychological test results in adults (Iverson & Binder, 2000). Poor effort can have a much greater effect on test scores than severe brain injury (Green, Rohling, Lees-Haley, & Allen, 2001). Therefore, effort tests are now a routine part of adult assessment but there is very little published evidence on the way in which children perform on such tests.

The assumption that all children will exert optimal levels of effort and motivation on testing has been challenged by several studies (Hart, 1995; McCann, 1998; Schmitz & Skinner, 1993). It has been shown that children can fake neuropsychological deficits when coached to do so and that the invalid data resulting from incomplete effort are not easily detected by clinicians (Faust, Hart, Guilmette, & Arkes, 1988a, 1988b; McCaffrey & Lynch, 1992; Rogers, 1997). Bigler (1990, p. 246) questioned the methods used in some of the latter work but acknowledged the need for further research. It is important to make it clear that poor effort in children does not necessarily imply deliberate faking of impairment. The fact that some children have little motivation to do well is reflected in the classical teacher's comment on children's report cards "Could do better". Children with ADHD often perform poorly on a variety of academic tasks and

Address correspondence to: Dr. Paul Green, Neurobehavioural Associates, Edmonton, Alta., Canada. Tel.: +1-780-484-5550. Fax: +1-780-484-5631. E-mail: paulgreen@shaw.ca
*Tel.: +1-780-483-8590. Fax: +1-780-455-6717. E-mail: drflaro@shaw.ca
Accepted for publication: February 27, 2003.

demonstrate significant motivation difficulties, even though they appear to have adequate cognitive capabilities (Barkley, 1999; Douglas, 1983; Hoza, Pelham, Waschbusch, Kipp, & Owens, 2001). In some cases, this could indicate a lack of consistent effort, yielding unreliable test data, even though the child does not necessarily intend to appear impaired. Lack of motivation is not the same as motivation to do poorly. In others, unreliable data could result from inability to sustain attention or a tendency to give up and not try when facing difficulty. However, if we do not measure effort objectively, it will be impossible to differentiate between attentional problems and poor or fluctuating effort.

The current study employed the Word Memory Test (WMT; Green, Allen, & Astner, 1996), which consists of both measures of effort and measures of verbal memory. The task involves learning a list of 20 word pairs, which is presented twice on the computer screen (or, optionally, read aloud by the examiner in the oral version). The effort measures are defined as those which are extremely easy for people to pass and on which the mean scores from adults with severe brain injury or neurological diseases, such as brain tumors, are about 95% correct (Allen & Green, 1999; Green & Allen, 1999; Green et al., 1996). Unlike most neuropsychological tests, which are ability measures (e.g., problem-solving, reading or commonly used memory tests), the scores on these effort subtests in adults are unrelated to age, intelligence and years of education (Green, Lees-Haley, & Allen, 2002). In contrast, the memory subtests, including a multiple choice (MC) task, a paired associates (PA) task and delayed free recall (DFR), are those which are affected by ability and which do produce lower than normal scores in patients with neurological impairment. Any test score may be affected by effort but some measures are far more sensitive to impaired ability than others. Effort measures should be affected very little by actual impairment of ability and, ideally, they should not be affected at all.

Hartman (2002) describes the WMT as one of the most widely researched and most sensitive tests of effort used with adults. The WMT recognition memory subtests have been validated as effort tests in a number of studies. It has been shown that most patients making an effort can easily score above the published cut-offs, including those with head injuries (Allen & Green, 1999; Green, Iverson, & Allen, 1999; Iverson, Green, & Gervais, 1999) fibromyalgia (Gervais et al., 2001) chronic pain syndrome (Gervais, Green, Allen, & Iverson, 2001), neurological diseases affecting the brain and various other diagnoses (Green & Allen, 1999; Green et al., 2001). However, in principle, what is easy for an adult will not necessarily be easy for a child. In some children, reading difficulty could interfere with performance on the WMT, if it were administered in the same way in which it is used with adults (i.e., no assistance with reading the stimulus items presented on the computer screen). To reduce the impact of reading problems, one option might be to administer the original oral form of the WMT, which would remove the need to read any words (Green et al., 1996). In the current study, the computerized administration was used but, as the study progressed, we realized that reading seemed to be a problem for some children whose reading was poor. For the remainder of the children, when their reading level was below grade 3, we tried to minimize the effects of reading difficulty on the WMT effort test scores, as explained in the Procedures section.

Flaro and Green (2000) reported the findings from a pilot study involving clinical testing of 69 children with the WMT. Only 10% of children failed the WMT effort subtests using the adult criteria defined in the test manual and those children who failed the WMT effort subtests also produced consistently lowered scores on independent memory tests. The children who failed the WMT effort tests had no monetary incentive to do poorly on testing but they freely admitted that they had made little effort to obtain high scores and that they could have done better. The reasons given for not trying their best were usually idiosyncratic, for example, saying that the tests reminded them of school and that they did not like school. These results highlighted non-monetary factors which could lead to unsatisfactory effort during testing and suggested the need for further validation of the WMT for use with children and adolescents.

The current study extended the latter pilot project by continuing to use the WMT with all consecutive children seen clinically, leading to results from a total of 135 children, aged 7–18 years, (i.e., the original 69 cases plus 66 new cases), allowing different diagnostic subgroups to be compared with each other. The children's WMT results were compared with those from 29 adults, who were assumed to be well motivated to score highly because they were seeking custody of their children through the courts. They were also compared with those from adults with either a mild head injury, a moderate to severe traumatic brain injury, or a neurological disease affecting the brain.

In adults, several methods have been used to show that the WMT effort subtests actually measure effort and we tried to reproduce these methods with children. For example, adult patients were easily able to score above the published cut-offs on these measures despite having severe brain injuries, strokes, brain tumors, ruptured aneurysms or other neurological diseases (Allen & Green, 1999; Green & Allen, 1999). This is the principle of insensitivity to actual impairment. In the current study, we examined the performance of children with clinically significant impairment, for example, including children with schizophrenia, neurological diseases, FAS or ADHD and there were some children of very low intelligence. Hypothesis A was that, consistent with the WMT effort subtests being relatively insensitive to actual impairment, children with probable cognitive impairment would have little difficulty passing these subtests using the adult criteria (Iverson et al., 1999).

An important and unusual method has been used with adults to demonstrate convincingly that the WMT effort subtests are relatively insensitive to ability but that, instead, they measure effort. The method involves showing an absence or even a reversal of the expected differences in WMT scores between two or more groups, which are expected to differ from each other in ability. For example, in the study of Green et al. (1999), the mild head injury group would have been expected to score higher on any ability measure than the definite brain injury group, which had well established moderate to severe brain injuries. Such a prediction would be consistent, for example, with the finding of Dikmen, Machamer, Winn, and Temkin (1995) that patients with relatively mild traumatic brain injury and shorter times to follow commands, scored higher on their test battery than did patients with more severe injuries and longer times to follow commands. Contrary to this expectation, the WMT effort scores were actually significantly higher in the definite brain injury group than in the mild head injury group. Such results could not be explained on the basis of differences in ability because more severe brain injuries should cause greater impairment than less severe injuries. The only reasonable explanation was in terms of lower effort or greater exaggeration of difficulty in the least severely injured patients.

The same principle of showing an absence of expected differences between patient groups, who are known to differ from each other in ability, may also be used to demonstrate that the WMT effort subtests measure effort in children, rather than mainly being measures of ability. In the current study, we examined the scores of children at various ages and at various levels of intelligence. If the WMT recognition subtests measured memory ability, we would expect to observe an increase in the children's scores with age. Similarly, if the WMT effort subtests were measuring ability, we would expect the children of this study to score lower on WMT effort subtests than a group of community dwelling parents. The parents were seeking custody of their children through the courts, whereas the children had significant psychiatric, neurological, behavioral and learning disorders. Also, because mild head injuries do not cause significant neuropsychological impairment relative to adult orthopedic control subjects (Dikmen et al., 1995) we would expect patients with mild head injuries to score higher than the children in the current study, many of whom had disorders known to be associated with cognitive impairment. Hypothesis B was, therefore, that, if they were measuring ability, (a) there would be an age effect on WMT Immediate Recognition (IR) and Delayed Recognition (DR) within the group of children, (b) that lower scores would be expected in children than in parents and (c) that lower scores would be expected in children than in adults with mild head injuries. Similarly, if the WMT recognition subtests were sensitive to differences in levels of ability, we would expect children of lower IQ to perform less well on WMT effort subtests than children of higher IQ. Hypothesis C was that there would be an IQ effect on WMT IR and DR.

The list of word pairs on the WMT is shown on a computer screen and so learning the list requires some reading skill. The recognition and multiple choice trials in the computerized WMT also involve reading words on a computer screen. Hypothesis D was that reading level would affect scores on both the WMT effort and memory subtests. It is to be expected that a certain minimum level of reading must be attained before a young child can score above the adult cut-offs on. For example, the average two-year old would obviously not be able to pass the WMT. However, before this study was done, it was not known what the minimum required reading level would be. Part of the way through the study, the examiner (second author) suspected that children with less than a grade 3 reading level were having trouble with the test because of reading problems. For this reason, the remaining children with less than a grade 3 reading level were given assistance with reading, using the method described in the next section. Hypothesis E was that children with a reading level of less than grade 3, who were given assistance with reading on the computerized WMT, would score higher on all WMT subtests than those of a similar reading level who were tested without any reading assistance. It has been shown that, in adults who failed the WMT effort subtests, scores on neuropsychological tests and especially on memory tests were markedly suppressed, compared with the scores from those who passed the WMT effort tests. This effect was so strong that patients with very mild head injuries who failed the WMT produced far lower scores on memory and other tests than patients with very severe brain injuries, who passed the effort subtests (Green et al., 2001). Thus, the WMT appeared to measure a disposition towards underachievement, which affected test scores in general, although memory test scores were affected more than non-memory test scores. Similarly, if the WMT were measuring incomplete effort in children, we would expect WMT failure to be linked with a generalized suppression of test scores. Therefore, the children in the current study were divided into those who passed or failed the effort tests of the WMT, using the adult cut-offs. These two groups were then compared in terms of their scores on independent memory tests, the Children's Auditory Verbal Learning Test (CAVLT; Talley, 1993) and the Rey Complex Figure Test (RCFT; Meyers & Meyers, 1996). Hypothesis F was that children who failed the WMT recognition subtests using the adult criteria would score lower than the remaining children on the independent verbal and visual memory tests.

## METHOD

### Participants

One-hundred and thirty-five children between the ages of 7 and 18 years participated in this study (mean age = 12.6 years, $SD = 2.7$). All children were referred to the second author's private practice for neuropsychological assessment by social services, school boards and physicians. The children were placed into broad diagnostic groupings, which were not designed to separate "functional" from brain-based pathology, but only to differentiate broad diagnostic categories. There were 19 children with FAS, which includes 7 with fetal alcohol effects for the purposes of this study. The "neurological disorder" group ($n = 16$) consisted of children with Wilson's disease, meningitis, myotonic dystrophy, cerebral palsy, pervasive developmental disorder, seizure disorder/mesial temporal sclerosis, Tourette's syndrome, Asperger's syndrome, fragile X syndrome and Leri–Weill syndrome. Another group consisted of children with psychiatric disorders ($n = 10$), including three cases of schizophrenia, six with bipolar mood disorder and one case of post traumatic stress disorder. Other diagnoses included ADHD ($n = 21$); learning disabilities ($n = 28$) and behavioral disorders, predominantly conduct disorder or oppositional-defiant disorder ($n = 41$). All children were given a comprehensive neuropsychological evaluation, including many tests but only the CAVLT, the RCFT and the WMT were given to all cases and so these are the only tests considered in this study.

### Procedures

In the computerized WMT, the subject is presented with 20 semantically linked simple word pairs (e.g., "pig–bacon"), each pair appearing on the screen for 6 s. After the list is presented twice, there is an immediate recognition (IR) trial, in which the person is shown new word pairs containing only one of the words from the original list and must select the words belonging to the original list (e.g., "pig" from the pair "cow–pig"). Without advance warning, a similar delayed recognition (DR) procedure is administered after 30 min, using different foil words paired with the original words (e.g., "feed – pig"). WMT scoring includes a calculation of consistency of responding from IR to DR (called "Consistency" in the tables below), and this measure is also sensitive to patient effort. The effort measures are followed on the WMT by a series of memory tests of gradually increasing difficulty. These include multiple choice (MC), in which the person is shown the first word of each pair and has to select the other word from eight options; paired associates (PA), in which the tester tells the person the first word and asks for the second word in each pair; delayed free recall (DFR), involving the person saying as many words as possible from the list in any order, while the tester records responses on the computer. The long delayed free recall (LDFR) subtest, which is the same as DFR but after another 20-min delay, was not used in the current study.

The recognition subtests (IR and DR) serve a dual function as effort measures and also as additional learning trials because all words from the list are presented and feedback is given on correctness of responses. The MC and PA trials provide further exposure to the 20 first words from the pairs, again serving a dual purpose as tests of memory and as learning trials, prior to free recall. Thus, as used in the current study, the WMT generated three effort scores and three memory scores.

When taking the computerized WMT, using the method described in the test manual (Green et al., 1996), people are given assistance with reading and understanding the task instructions if needed, but they receive no assistance with reading the stimulus items presented on the computer screen. For various reasons, if a person is expected to have difficulty taking the computerized test (e.g., blindness, problems with English fluency, severe reading difficulty, advanced dementia or Korsakoff's psychosis), use may be made of the oral WMT, which requires no reading, except on the multiple choice trial.

In order to compare the children with a group of community dwelling adults, who were presumed to be making a good effort, 29 adults were selected (18 women, 11 men). These people were parents, who were seeking custody of their children. They were undergoing a psychological assessment to assist the courts to determine whether they would be granted sole or joint custody of their children. The parents' mean educational level was 11 years ($SD = 2.3$) and their mean age and intelligence scores are shown in Table 1. The parents were unrelated to the children in this study.

It is important to compare the scores of children on effort tests with those from adult patients. For this reason, we included adult patients with head injuries or a variety of neurological conditions including patients described previously in the study of Green et al. (2001) with the addition of patients since evaluated as part of the same consecutive series. Patients with head injuries were divided into two groups. The moderate–severe group ($n = 104$; Table 2) had posttraumatic amnesia (PTA) of 1 day or more (median PTA duration = 336 hr). Their median Glasgow Coma Scale score was 8 and all were selected for having brain abnormalities on CT or MRI scans. Their mean education was 11.6 years ($SD = 2.5$) and 89% were men. The other group, with mild head injuries ($n = 206$) had less than 1 day of PTA (median duration of PTA = 0 hr). Their median Glasgow Coma Scale score was 15 and they were all selected for having no abnormality on a CT or MRI brain scan. Their mean education was 11.9 years ($SD = 2.9$) and 75% were men.

A final group of neurological patients ($n = 104$) suffered from a variety of disorders, including 36 strokes, 15 aneurysms, 14 multiple sclerosis, 14 brain tumor, 4 epilepsy and 26 cases with miscellaneous conditions (e.g., herpes simplex encephalitis, Von Hippel-Lindau disease, hypoxic event, abscess, venous thrombosis, dorsal midbrain hemorrhage). In the neurological patients, 92% of cases had intracranial abnormalities on CT or MRI scans. Their mean age was 47.6 years ($SD = 9.5$) and they had a mean of 13.2 years of education ($SD = 3.4$). All adult patients were seen for neuropsychological assessment as outpatients in the context of a Canadian Workers' Compensation Board claim, a medical disability claim or personal injury litigation.

All children and adults in this study were given the computerized WMT. For 107 children with at least a grade 3 reading level on the Test of Academic Performance (Adams, Sheslow, & Erb, 1989), the adult administration was used, in which the person reads the list of word pairs silently as it appears on the screen and later selects the target words from word pairs presented on the computer screen. For the first 13 cases with less than a grade 3 reading level, no special assistance was given. In the course of the study, the examiner (the second author) judged that these children were having some difficulty with reading the WMT stimuli. For the next 14 cases with less than a grade 3 reading level, assistance with reading was given. The child was asked to read aloud each word as it appeared on the computer screen in the two initial learning trials. If the child had any difficulty, the tester said the word aloud immediately and this was done on both presentations of the list of word pairs. Also the examiner asked the child to read out each of the word pairs presented on the immediate and delayed recognition trials and, if an error was

Table 1. Mean Scores in Percent Correct (and Standard Deviations) on WMT Subtests From Children in Each Diagnostic Group and in a Group of Parents.

| Group | Demographics | | | WMT effort scores | | | WMT memory scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Age | VIQ | IR | DR | IR–DR consist | MC | PA | DFR |
| (1) ADHD | 21 | 10.9 (2.7) | 93.0 (14.7) | 94.0 (6.3) | 95.1 (7.3) | 91.8 (8.3) | 86.4 (14.8) | 84.0 (16.8) | 44.0 (17.4) |
| (2) Conduct Disorder | 41 | 14.0 (2.4) | 90.3 (13.9) | 95.4 (6.6) | 96.3 (6.8) | 93.8 (8.0) | 85.4 (20.7) | 85.9 (16.3) | 49.4 (14.4) |
| (3) FAS | 19 | 12.4 (2.7) | 88.8 (12.8) | 96.1 (4.4) | 95.5 (5.8) | 94.7 (5.9) | 77.6 (29.0) | 81.1 (26.9) | 45.9 (23.2) |
| (4) Learning Disabilities | 28 | 12.3 (2.6) | 88.8 (15.0) | 94.6 (6.0) | 94.0 (9.0) | 90.6 (9.5) | 83.4 (17.9) | 81.4 (17.5) | 44.9 (17.7) |
| (5) Neurological Disorders (children) | 16 | 11.7 (2.2) | 94.3 (18.8) | 94.5 (9.6) | 93.8 (11.1) | 91.4 (12.6) | 82.9 (21.1) | 77.3 (26.5) | 41.5 (20.7) |
| (6) Psychiatric Disorders | 10 | 13.4 (3.3) | 90.7 (24.4) | 98.4 (2.0) | 99.1 (1.7) | 97.5 (3.2) | 89.1 (17.4) | 84.1 (20.7) | 52.7 (26.4) |
| (7) Parents seeking custody | 29 | 34.2 (6.7) | 93.7 (15.8) | 96.8 (3.4) | 97.6 (4.3) | 95.2 (5.4) | 92.6 (11.9) | 87.4 (12.3) | 58.9 (17.1) |
| Significance | – | <.001 | ns | ns | ns | ns | ns | ns | <.03 |

*Note*. ADHD = attention deficit disorder with hyperactivity; FAS = fetal alcohol syndrome/fetal alcohol effects, 17 of 19 cases being of Native Indian descent; "Consist" = consistency between responses on IR and DR; VIQ = verbal IQ; IR and DR = immediate and delayed recognition; IR–DR consist = consistency of responses from IR to DR; MC = multiple choice; PA = paired associates; DFR = delayed free recall of word list.

Table 2. Mean Scores in Percent Correct (and Standard Deviations) on WMT Subtests From Children Compared With Adults With Relatively Mild or Moderate–Severe Brain Injuries and Neurological Diseases.

| Group | Demographics | | | WMT effort scores | | | WMT memory scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Age | VIQ | IR | DR | IR–DR consist | MC | PA | DFR |
| All children | 135 | 12.6 (2.7) | 90.5 (15.5) | 95.2 (6.4) | 95.5 (7.6) | 93.0 (8.6) | 83.9 (20.4) | 82.7 (19.9) | 46.3 (18.5) |
| Mild head injury (adults): median GCS = 15 | 197 | 39.5 (10.9) | 96.2 (13.0) | 85.5 (16.6) | 83.2 (17.4) | 81.4 (16.4) | 71.5 (24.0) | 66.9 (25.2) | 42.1 (17.6) |
| Mod-severe brain injury (adults): median GCS = 8 | 101 | 38.0 (12.6) | 93.7 (13.4) | 92.1 (9.9) | 92.1 (10.9) | 89.4 (10.4) | 77.5 (22.0) | 71.3 (24.0) | 40.6 (19.6) |
| Neurological diseases (adults) | 104 | 46.5 (9.3) | 99.7 (14.5) | 91.5 (13.7) | 91.1 (13.7) | 89.5 (11.9) | 78.1 (21.6) | 72.2 (23.6) | 41.8 (18.0) |
| Significance | – | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 |

*Note*. Paradoxically lower effort scores in mild versus severe brain injury indicate lower effort in the former. Parents seeking custody were also included in the above analyses but their data are shown in Table 1. For abbreviations, see footnote of Table 1.

made, the examiner said the correct words aloud but did not provide any prompt to aid selection of the correct response. If no errors were made on the first trial, the child was allowed to complete the test without any additional assistance. Otherwise, the administration was the same as with adults. No assistance was given with the performance of the remaining WMT subtests.

## RESULTS

For normally distributed data, such as verbal intelligence or WMT free recall scores, ANOVA was used to test for differences in group mean scores. Non-parametric tests (Mann–Whitney or Kruskall–Wallace) were used to compare groups on non-normally distributed data, such as WMT IR and DR effort test scores.

Hypothesis A (insensitivity to impairment of WMT IR and DR) was supported. The median scores on the WMT IR and DR "effort subtests" were at or above 97.5% correct in all six of the child diagnostic groups and in the custody-seeking parents. Table 1 shows the mean scores of the children in each diagnostic subgroup and of the custody-seeking parents on the three WMT effort measures, which are WMT IR (immediate recognition subtest), WMT DR (delayed recognition subtest) and consistency (consistency of responses from the IR to the DR subtest). Using a Kruskall–Wallace test, there was no significant difference between groups on IR, DR, or consistency between the child diagnostic groups and independently living adults seeking custody of their children ($p > .10$ for all variables).

Table 1 also shows the mean scores from each subgroup on those subtests of the WMT that were designed to measure memory ability instead of effort (i.e., MC, PA, and DFR). The parents scored significantly higher than all children combined on MC, $F(1, 162) = 4.8$, $p < .05$, and the DFR or delayed free recall subtest, $F(1, 162) = 11.3$, $p < .01$. When the parents' WMT memory subtest scores were compared with those from all child diagnostic groups in an ANOVA, the only significant difference between groups was on the most difficult subtest, the DFR or delayed free recall subtest, $F(6, 157) = 2.4$, $p < .05$. On the latter recall subtest, the parents obtained a mean score of 58.9% correct ($SD = 17.1$), which is close to the published mean of 63.7% ($SD = 12.4$) from adult volunteers of bright normal intelligence (Iverson et al., 1999). In contrast, the child neurological group mean was 41.5% ($SD = 20.7$) and significantly lower than that of the parents ($p < .05$ on post hoc bonferroni comparison).

Contrary to expectation, two of the parents showed poor effort on the WMT by scoring at or below 82.5% on one or more of the effort measures (IR, DR or consistency). They both admitted that they had not made an effort on testing and said that they had changed their minds and no longer wanted custody of their children. The mean effort scores in these two cases of self-admitted incomplete effort were IR = 88.7%, DR = 85% and consistency = 78.7%. The mean consistency score is below the adult cut-off for poor effort but the mean IR and DR scores are above the cut-off. This illustrates the fact that the cut-offs were set very conservatively, so as to minimize false positive identifications of poor effort in those who are truly very severely impaired (Green & Allen, 1999) but it also shows that the WMT was accurate in identifying inadequate effort in these two cases. The memory subtest scores in the two adults with poor effort were MC = 60%, PA = 62.5% and DFR = 33.7%.

Table 2 shows results from children in the current study compared with patients with mild head injuries versus moderate to severe brain injuries and a group of patients with various other neurological conditions. As previously argued by Green et al. (2001), the higher scores in the moderate-severe brain injury group compared with the mild head injury group are explainable by lower effort, on average, in the mild group. On all WMT effort measures, there were statistically significant differences between the groups in Table 2, with the combined children's group scoring the highest and the mild head injury group scoring the lowest.

To examine these group differences further, we used the adult criteria to define those who passed or failed the WMT effort measures. Each case was classified as a "pass" if IR, DR and consistency were all above 82.5% and, if not, the person was classified as "failing" the WMT effort tests.

The difference in failure rates between groups was statistically significant (chi square = 31, $df = 3$, $p < .01$). On the WMT effort tests, the failure rate of the children in this study was 14%. In contrast, in the adult patient groups assessed in the context of a compensation claim, the WMT effort test failure rates were as follows: mild head injury group, 45%; moderate-severe brain injury group, 23%; neurological group, 18%. WMT effort test failures in the children tested clinically were clearly far less prevalent than in the adult patient groups. In the adult patients passing the WMT, 56%, had abnormal findings in CT or MRI brain scans, whereas only 28% of those failing the WMT had abnormalities on brain scans.

Table 3 shows that, on WMT IR, the children who passed the WMT scored no differently from the adult patients who passed the WMT. In those passing the WMT, the children scored slightly but significantly higher than the adults on WMT DR (Mann–Whitney, $z = -3.1$, $p < .01$), on consistency ($z = 2.2$, $p < .05$) and on Paired Associates ($z = 2.1$, $p < .05$) but the differences were non-significant on the remaining WMT subtests.

In those who failed the WMT effort subtests, it is notable that the children scored significantly higher than the adult patient groups on all WMT effort measures (IR, $z = 2.24$, $p < .05$; DR, $z = 2.6$, $p < .01$; consistency, $z = 2.4$, $p < .05$) and also on the Paired Associate subtest, $F(1, 142) = 18.8$, $p < .01$. They did not differ significantly on WMT Multiple Choice or Delayed Free Recall subtests. Thus, there were no examples of the adult patients scoring higher than the children on any WMT subtest.

Hypothesis B (that there would be a significant age effect on the effort measures) was rejected. Firstly, as shown above, there was no superiority on WMT effort subtests in the adults compared with the children. The 135 children did not score any lower on the WMT effort subtests than the parent group and this is clearly contrary to what would be expected if there were a major age effect in people aged seven years or more. Table 4 also shows the mean scores on all WMT subtests in children divided into different age levels (means of 8, 11, 14 and 16 years of age). There were no statistically significant differences between these age groups on WMT effort measures ($p > .10$). It should be noted that the table shows statistically significant differences between age groups in reading levels ($p < .01$), which varied from a mean of grade 2 in the youngest group to a mean of 8 in the oldest group. Although the WMT IR differences between age levels were non-significant, lower reading levels might account for the fact that there was a trend towards lower effort scores in the youngest group. For example, the mean WMT IR score was 91.7% ($SD = 11.7$) in the youngest group but it was stable at 95–96% in all other age groups. However, when reading level was controlled in a partial correlation, there was no statistically significant correlation between age and any of the WMT effort measures. We also controlled for the effect of reading by selecting children with reading levels between grades 3 and 8, which removed significant differences in reading levels between the age groups. After this procedure, the mean WMT IR score in the youngest group (95.5%, $SD = 3.2$, $n = 5$) was not different from the mean scores in the other age groups, although the sample is small.

As we would expect from tests that are sensitive to differences in ability, the older children scored significantly higher than the younger children on the WMT memory ability subtests, including MC, $F(1, 133) = 12.2$, $p < .01$; PA, $F(1, 133) = 22.2$, $p < .01$ and DFR, $F(1, 133) = 34.9$, $p < .01$.

Hypothesis C, that there would be an IQ effect on WMT IR and DR, was rejected. The WMT is a verbal task and so verbal intelligence (VIQ) is most directly relevant to expected performance on this test. Table 5 shows the WMT scores from children broken down into four different ranges of verbal intelligence, the mean VIQ scores per group being, respectively, 64, 76, 91 and 110. Despite the major differences in VIQ between groups, there were no statistically significant differences between them on the effort measures of the WMT ($p > .10$). When the children were broken down into two groups, those with a VIQ of 70 or below (mean VIQ = 64.0, $SD = 4.5$) versus those with a VIQ greater than 70 ($M = 93.3$, $SD = 13.5$), there were still no statistically significant differences between these VIQ groups. There was no statistically significant correlation

Table 3. Mean Scores in Percent Correct (and Standard Deviations) on WMT Subtests in Those Who Passed or Failed WMT Effort Tests, Using the Adult Cut-offs: Children Versus Combined Adults With Brain Injuries and Neurological Patients.

| Group | Demographics | | | WMT effort scores | | | WMT memory scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Age | VIQ | IR | DR | IR–DR consist | MC | PA | DFR |
| Children *passing* effort subtests | 116 | 12.6 (2.7) | 91.5 (15.5) | 97.1 (3.7) | 97.8 (3.3) | 95.7 (4.8) | 87.8 (16.7) | 85.2 (17.6) | 48.6 (18.1) |
| Adult patients *passing* effort subtests | 263 | 40.9 (11.8) | 99.0 (13.5) | 96.6 (3.7) | 96.2 (3.9) | 94.3 (4.8) | 87.1 (12.8) | 81.8 (16.5) | 48.7 (16.2) |
| Children *failing* effort subtests | 19 | 12.5 (3.0) | 84.6[a] (13.9) | 84.1 (7.9) | 81.2 (10.8) | 76.3 (8.4) | 60.1 (24.9) | 67.1[a] (26.1) | 31.7[a] (14.2) |
| Adults *failing* effort subtests | 137 | 41.8 (11.3) | 91.2[a] (12.7) | 73.5 (16.2) | 70.6 (15.9) | 68.5 (11.6) | 49.9 (18.5) | 44.5[a] (18.5) | 27.2[a] (12.8) |

[a]Note that scores on any test, such as VIQ, WMT DFR or MC can be affected by poor effort. For abbreviations, see footnote of Table 1.

Table 4. Mean Scores in Percent Correct (and Standard Deviations) on WMT Subtests From Children in Four Age Groups.

| Group | Demographics | | | WMT effort scores | | | WMT memory scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | Reading grade | Age | VIQ | IR | DR | IR–DR consist | MC | PA | DFR |
| 7–9 years ($n = 22$) | 2.7 (1.4) | 8.2 (0.7) | 87.3 (14.0) | 91.7 (8.7) | 91.6 (11.7) | 89.0 (10.6) | 66.4 (25.8) | 61.4 (24.9) | 25.7 (13.7) |
| 10–12 years ($n = 39$) | 5.5 (2.2) | 11.2 (0.8) | 92.3 (17.4) | 96.0 (4.6) | 96.8 (5.2) | 94.1 (7.1) | 89.5 (12.1) | 88.4 (12.7) | 46.1 (16.1) |
| 13–15 years ($n = 55$) | 7.4 (2.7) | 14.0 (0.8) | 90.1 (14.5) | 96.0 (5.8) | 95.8 (6.2) | 93.6 (8.3) | 88.8 (17.2) | 85.6 (18.2) | 51.7 (15.2) |
| 15–18 years ($n = 19$) | 8.5 (2.7) | 16.7 (0.7) | 91.9 (16.1) | 95.2 (7.0) | 96.0 (8.8) | 93.4 (9.1) | 78.6 (23.9) | 87.4 (14.1) | 54.4 (20.5) |
| Significance | <.001 | <.001 | ns | ns | ns | ns | <.001 | <.001 | <.001 |

*Note*. In all tables, non-parametric statistics were used for group comparisons using non-normal data (i.e., IR, DR and consistency). For abbreviations, see footnote of Table 1.

Table 5. Mean Scores in Percent Correct (and Standard Deviations) on WMT Subtests From Children of Differing Levels of Verbal Intelligence.

| VIQ range | Demographics | | | WMT effort scores | | | WMT memory scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Age | VIQ | IR | DR | IR–DR consist | MC | PA | DFR |
| (1) 70 or lower | 13 | 11.7 (2.7) | 64.0 (4.5) | 90.9 (15.7) | 87.5 (25.0) | 88.1 (12.9) | 56.9 (31.8) | 56.9 (25.3) | 25.0 (14.9) |
| (2) 71–80 | 24 | 12.8 (2.6) | 75.9 (2.6) | 94.4 (7.1) | 95.9 (6.4) | 91.4 (9.3) | 86.6 (12.1) | 77.1 (18.4) | 43.8 (13.5) |
| (3) 81–100 | 64 | 12.8 (2.8) | 90.8 (5.8) | 95.4 (6.5) | 95.5 (8.1) | 93.7 (8.2) | 85.4 (19.8) | 86.9 (17.6) | 48.8 (18.6) |
| (4) 101 or higher | 34 | 12.4 (2.6) | 110.5 (7.7) | 95.5 (6.6) | 95.4 (8.1) | 93.5 (8.9) | 86.9 (19.5) | 86.9 (17.10) | 51.2 (17.5) |
| Significance | – | ns | <.001 | ns | ns | ns | <.001 | <.001 | <.001 |

*Note*. For abbreviations, see footnote of Table 1.

between VIQ and either WMT IR or WMT DR in the children or in the parent group. The absence of significant effects of VIQ on WMT IR and DR in children adds further support to the suggestion that these effort measures are quite insensitive to differences in ability levels in the children of this study. In contrast, as we would expect, there were highly significant differences between children of different VIQ levels on all three WMT memory subtests, including MC, $F(3, 131) = 8.3$, $p < .01$; PA, $F(3, 131) = 10.9$, $p < .01$ and DFR, $F(3, 131) = 8.1$, $p < .01$ (Table 5).

In order to study the WMT IR and DR scores in young children of low verbal intelligence, children were selected who were less than 11 years of age and who had a VIQ of less than 75. There were 7 such children. Their mean VIQ was 68 ($SD = 4.5$), their mean age was 8.8 years ($SD = 1.1$) and their mean reading grade level was 2.3 ($SD = 1.2$). Their mean WMT IR score was 95% ($SD = 4.5$), their mean WMT DR score was 98.2% ($SD = 2.4$) and their mean consistency score was 93.5% ($SD = 5.5$).

Hypothesis D, that there would be a reading effect on WMT IR and DR scores, was accepted. Reading grade levels on the Test of Academic Performance (Adams et al., 1989) were available for 134 children. One 16-year-old boy refused reading testing. When age and verbal intelligence were controlled in partial correlations, reading grade level correlated significantly with WMT IR ($r = 0.29$, $p < .01$), WMT DR ($r = 0.23$, $p < .01$) and consistency ($r = 0.27$, $p < .01$). In Table 6, the children with the lowest reading level ($M =$ grade 1.7) were, on average, 10-year old with a mean VIQ of 81. They scored a mean of 88.6% correct ($SD = 10$) on WMT IR and 86.7% correct ($SD = 12.7$) on WMT DR, whereas children in groups with higher reading levels scored significantly higher on IR and DR, with mean scores between 94% and 97% correct ($p < .01$). Thus, a reading level below grade 2 was clearly associated with a statistically significant drop in mean WMT IR and DR (effort) scores of about 6–10% points relative to other children.

There was a statistically significant difference in failure rates on WMT effort tests between the groups of children of different reading grade levels shown in table 6 (Chi Square = 13.0, $df = 4$, $p < .01$). Forty-six percent (i.e., 6/13) of the children with a grade 2 reading level or below ($M = 1.7$, $SD = 0.3$) failed the WMT by the adult criteria, whereas the failure rate was only 10.7% (13/121) in all the other children, who had greater than a grade 2 reading level ($M = 6.7$, $SD = 2.7$; $z = 3.4$, $p < .01$). When children with a grade 2 reading level or below were dropped from the analysis, there was no significant difference in rates of failure on the WMT between the remaining groups shown in Table 6 (Chi Square = 1.2, $df = 3$, $p > .10$). The failure rates were 11%, 9%, 5%, and 15%. That is, 11% of cases (5/44) with a reading grade level of 8 or higher failed the WMT effort subtests; 9% (3/32 cases) failed the WMT in reading grades 6.0–7.9; 5% of cases (1/19) failed the WMT in reading grades 4.0–5.9 and 15% of cases (4/26) failed the WMT in reading grades 2.1–3.9. When the lowest reading level group in Table 6 was dropped, leaving 116 children, there were no significant differences between the remaining groups on WMT IR, WMT DR or consistency despite the wide range of reading levels. These data suggest that the critical reading level below which reading problems cause children to have difficulty with the WMT is approximately grade 3.

The computerized test requires reading and so it is not surprising that reading levels correlated highly with scores on the memory subtests, including Multiple Choice ($r = 0.4$), Paired Associate Recall ($r = 0.51$) and the Delayed Free Recall subtest ($r = 0.61$, all $p < .01$). In contrast, the CAVLT involves hearing the word list read aloud by the tester and it requires no reading. Correspondingly, the correlations between reading levels and scores on the CAVLT were weaker, although statistically significant at $r = 0.21$ (immediate recall) and $r = 0.28$ (delayed recall).

Hypothesis E was that reading assistance would lead to an increase in WMT scores in those with reading levels less than grade 3. Contrary to expectation, Hypothesis E was not supported. There were no differences in mean scores on any of the WMT subtests between 14 children with low reading levels who were given assistance with reading (Group A) versus the 13 given no such assistance (Group B). These two groups also

Table 6. Mean Scores in Percent Correct (and Standard Deviations) on WMT Subtests From Children of Differing Reading Levels.

| Reading grade | Demographics | | | WMT effort scores | | | WMT memory scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Age | VIQ | IR | DR | IR–DR consist | MC | PA | DFR |
| 1.7 (0.3) | 13 | 10.3 (2.7) | 81.1 (15.9) | 88.6 (10.0) | 86.7 (12.7) | 83.4 (13.5) | 62.1 (22.4) | 55.4 (22.9) | 23.2 (15.9) |
| 2.9 (0.6) | 26 | 10.5 (2.6) | 80.6 (12.1) | 94.1 (6.3) | 94.5 (8.2) | 91.1 (8.3) | 76.1 (24.4) | 71.7 (21.3) | 34.4 (13.6) |
| 5.0 (0.6) | 19 | 11.9 (2.0) | 87.3 (16.6) | 96.7 (3.5) | 97.1 (5.9) | 94.6 (7.3) | 86.6 (20.1) | 88.4 (20.1) | 45.9 (15.5) |
| 6.9 (0.7) | 32 | 13.0 (2.3) | 91.4 (12.2) | 96.1 (4.6) | 97.8 (3.2) | 94.5 (5.6) | 89.5 (11.0) | 87.5 (12.1) | 48.4 (12.9) |
| 9.7 (1.2) | 44 | 14.5 (1.8) | 99.8 (13.7) | 96.5 (6.1) | 96.1 (6.7) | 94.9 (7.7) | 89.7 (17.7) | 91.4 (12.7) | 58.6 (16.4) |
| Significance | – | <.001 | <.001 | <.01 | <.01 | <.01 | <.001 | <.001 | <.001 |

did not differ from each other in terms of their mean age (Group A = 10 years, $SD = 2.5$; Group $B = 9.6$, $SD = 2.5$), VIQ (Group A: VIQ = 82.7, $SD = 13.7$; Group B: VIQ = 81.7, $SD = 14.0$) or reading level (Group A: reading = grade 2.1, $SD = 0.4$; Group B: reading = grade 2.1, $SD = 0.5$). There was no significant difference between the rates of failure on the WMT in the assisted versus unassisted groups (4 vs. 3, respectively).

Having a specific reading disability does not, in itself, imply difficulty with the WMT, as long as the child's reading level is above grade 2. For example, a group of children with reading difficulty was selected as having (a) at least an IQ of 85 on both verbal and performance scales and (b) having a reading level more than 1 year behind educational level ($M = 2.3$ years, $SD = 0.8$) but (c) having a reading level of 2.3 or higher. (Note: If we wish to compare readers of higher versus lower abilities by dividing a group into two, a cut-off score must be chosen. Any score that divides one part of the group from the other is arbitrary, whether it is an exact year like grade 3.0 or part of a year, like grade 3.2.) There were 14 children in the reading difficulty group meeting criteria "a" to "c" above. Their mean years of education was 7.8 ($SD = 1.1$); their mean age was 13.1 years ($SD = 2.1$), their mean reading grade level was 5.4 ($SD = 1.7$) and their mean verbal and performance IQ scores were, respectively, 93.5 ($SD = 6.5$) and 99.7 ($SD = 9.9$). Twelve out of fourteen of these children passed the WMT using the adult cut-offs. In contrast, if children were selected not only for a reading level more than 1 year behind their educational grade but also a grade 2.3 reading level or lower, 45.5% failed the WMT (5/111 cases). The difference between educational level and reading level was calculated and there was no correlation between this variable and any of the WMT effort subtests. This result was also found both when all children with less than a grade 2.3 reading level were dropped and also when reading level was controlled in a partial correlation. The results suggest that having a specific reading difficulty relative to age and intelligence does not lead to WMT failure, as long as reading is above a certain level. For practical purposes, the results suggest that a grade 3 reading level may be required to pass the WMT effort tests.

Hypothesis F (generalization of poor effort) was accepted. As shown in Table 3, there were 19 out of 135 children (14%) who "failed" the WMT effort tests using the adult cut-offs (less than or equal to 82.5% on IR, DR or consistency). Five of them were just at the cut-off and scored no lower than 82.5% on any of the latter measures but they were still classified as "failures". The mean reading level of these children was grade 4.7 ($SD = 2.9$). On the WMT memory subtests, those that failed effort criteria scored between 0.9 and 1.7 *SD* below the means from the children who passed the WMT effort subtests.

On an independent measure of verbal memory (CAVLT), the children who failed WMT effort subtests also scored significantly lower than those who passed those measures by about two thirds of a standard deviation. Those who failed the WMT scored a mean of 7.4 ($SD = 2.4$) on CAVLT immediate recall, representing a drop of 0.7 *SD* relative to the mean score in those who passed the WMT, which was 9.3 ($SD = 2.8$), $F(1, 131) = 7.2$, $p < .01$. The mean CAVLT delayed recall score was 8.7 ($SD = 2.8$) in 114 children who passed the WMT effort measures versus 6.9 ($SD = 2.8$) in the 19 children who failed them. This represents a drop of 0.65 *SD* in the CAVLT mean score in cases of possible poor effort, $F(1, 131) = 6.3$, $p < .05$ Similarly, on a visual memory test (RCFT), those who failed the WMT effort subtests scored significantly lower than the remaining children, both on immediate and delayed recall of the design. The mean RCFT immediate recall score was 16.9 ($SD = 7.4$) in 104 children passing the WMT but only 12.0 ($SD = 5.6$) in the 18 participants failing it, a drop of 0.66 *SD*, $F(1, 124) = 7.1$, $p < .01$. The corresponding mean delayed recall scores were 15.8 ($SD = 7.4$) versus 10.4 ($SD = 5.1$), a significant drop of 0.73 *SD* in the probable poor effort cases compared with those making a good effort on WMT, $F(1, 124) = 8.8$, $p < .01$. Thus, failing the effort measures of the WMT was associated with a significant general reduction in levels of performance, not only on the WMT memory subtests but also on other independent memory tests, both verbal and visual-spatial, suggesting generally lowered effort, at least across tests of memory. It may be noted that significant generalized effects of this type were not observed on non-memory tests in this study (e.g., Grooved Pegboard, Peabody Picture Vocabulary Test, Finger Tapping Test) but the number of children given such tests and failing the WMT were too small to test the generalization hypothesis.

We gently questioned children who, by the adult criteria, had "failed" the WMT effort tests, telling them that their test scores suggested that they were not trying their best. Those children who failed the WMT effort tests freely admitted that they had made a poor effort and said that, if they had tried, they could have done much better, not only on effort tests but also on other tests . One boy said "Believe it or not, when I was younger, I used to fake being sick". In order to test whether the children who could actually have done better, near the end of the study, six consecutive children who "failed" the WMT were given the option of taking the test again on the same day. With parental permission, if they "passed" the WMT the second time, they would receive a small prize (their choice of a can of pop or a candy from a shelf). Five of them who had initially "failed" the WMT went on to "pass" by scoring above the adult cut-offs on all WMT effort scores. The increases in their mean scores from test to retest were as follows. IR was 81.5% ($SD = 8.7$) and increased to 93% ($SD = 4.4$); DR increased from 79% ($SD = 13.2$) to 95% ($SD = 3.9$); consistency was 72.5% ($SD = 8.2$) initially and was 90% ($SD = 3.9$) on retest; MC went from 54% ($SD = 22.7$) to 83% ($SD = 11.5$); PA climbed from 48% ($SD = 16.8$) to 70% ($SD = 6.1$) and DFR increased from 27.5% ($SD = 3.0$) to 37% ($SD = 6.4$). One other child, whose initial consistency score was 82.5%, failed to follow instructions. When taking the test the second time, he showed precipitous declines in his scores from 92.5% on IR, 85% on DR and 85% on MC to, respectively, 5%, 7.5% and 0%. It is clear that this boy chose to perform even worse the second time because his IR and DR retest scores were significantly worse than chance and far worse than his initial scores. His diagnosis was "oppositional-defiant disorder".

If we leave the latter boy's data in the analysis but remove the data from the five children who passed the WMT on retesting, when given an incentive, the percentile ranks relative to the remaining 130 children would be as follows. A score of 82.5% would be at the fifth percentile for WMT IR and WMT DR but 82.5% would be at the 11th percentile for consistency. A score of 75% would be at the third percentile for WMT IR and DR and at the fifth percentile for consistency. In other words, the fifth percentiles would be 82.5% for WMT IR, 82.5% for WMT DR and 75% for consistency.

## DISCUSSION

Good effort measures should be as insensitive as possible to actual impairment. The more that an effort measure is affected by variations in ability, the less sensitive it will be as a measure of effort. The first hypothesis in the current study was that the WMT effort subtests would not be sensitive to impairment and it was supported by the results. The parents seeking custody of their children scored no higher on the WMT effort subtests than did the children with significant clinical disorders, such as FAS, ADHD and neurological diseases, indicating that these subtests are extremely easy, even for children with clinically significant disorders.

Another sign that the WMT effort scores measured effort rather than ability is that they were unrelated to age in these children, who were between 7 and 18 years of age, or to variations in verbal intelligence, despite the fact that the children's VIQ values ranged from 52 to 129. The child with a VIQ of 52 scored 100% correct on immediate and delayed recognitions (IR and DR) and, hence, 100% on consistency of responses from IR to DR. However, her multiple choice, paired associates and delayed free recall scores were 50%, 65% and 25%, respectively, all of which are well below the means for the children in this study. Hence, this 17-year-old girl, with a diagnosis of schizophrenia, displayed good effort but impaired memory, compared with the other children in this study or compared with normal adults (Iverson et al., 1999).

The children scored significantly higher on the effort measures than adults with the mild head injuries. Using the adult criteria to define failure on the WMT effort subtests, only 14% of the children in this study failed the WMT effort measures, compared with a failure rate of 18% in adult neurological patients with abnormal brain scans and 45% in compensation-seeking adults with mild head injuries and no brain scan abnormalities. In the children of the current study, those who failed the WMT effort measures scored significantly lower than the remainder of the group on two independent memory tests (CAVLT and RCFT). Hence, in these children, poor effort on the WMT appeared to have a generalized effect on other memory tests, as previously reported in adults (Green, Rohling et al., 2001).

There is evidence that the children who failed the WMT, using the adult cut-offs, could have done better if they had chosen to do so. In 5 out of 6 children who initially "failed" the WMT effort tests and who were retested with a small incentive, the retest scores were all above the WMT effort cut-offs and they were quite similar to those of the children as a whole. The only child who did not pass the WMT the second time failed it deliberately, as shown by scores that were significantly worse than chance. The retest results suggest that the children's low scores on WMT effort measures on first testing were a result of poor effort, as admitted by the children themselves on post test questioning. The fact that a small incentive seemed to make a large difference to test scores in these cases raises the question of how much fluctuating effort contributes to unreliability in children's test scores in clinical practice or in research studies. On the other hand, there were only six children offered the extra incentive on retest in this study and so further research will be needed to confirm this finding. Ideally, an extra group would be included to control for the effects of practice alone. This group would consist of children who fail WMT and are asked to take it again but who are given no feedback on performance and no extra incentive for "passing" on retest.

Both age and intelligence were found to be highly significantly related to scores on the memory ability subtests of the WMT, which were designed to measure memory for the word list and which were intended to be sensitive to verbal memory impairment (Multiple Choice, Paired Associates and Delayed Free Recall subtests). Yet, the IR, DR and consistency variables were not sensitive to differences in age or verbal IQ. This is a further sign that the latter subtests measure the amount of effort applied to the test.

The children with the lowest reading level scored lower than other children on the WMT effort measures. Thus, it was not age or intelligence that were of most importance in affecting WMT effort test scores in this study but reading level. More specifically, in children with a grade 2 reading level or below, the mean WMT effort scores were lower than in the remainder of children. Once the children could read at a grade 3 level, they appeared to be as capable as any others of performing at a very high level on the WMT effort subtests. In future studies of computerized WMT scores in children, it will be important to measure reading ability, especially word identification and decoding skills. Otherwise, we might be misled into concluding that either age or intelligence are significantly related to WMT effort scores, which is clearly not the case in the current study. Special care will be needed with children whose reading level is below grade 3. Contrary to expectation, assisting children with reading on the WMT, when their reading level was below grade 3, did not have any significant impact on WMT test scores in this study. In such cases, it would probably be desirable to use the oral version of WMT but the oral version is not yet normed in children.

The current findings do not provide normative data on the WMT from healthy children who are free of psychiatric or neurological disorders. Such data will be needed, especially if the WMT is to be used to measure children's abilities to learn and remember verbal information which they have read. However, the results of this study suggest that, in the same way that the WMT effort subtest scores in adults are unaffected by variables such as age and intelligence, they are also very insensitive to variations in age and intelligence in children. The current findings suggest that even serious illnesses such as schizophrenia, FAS or attentional disorders, such as ADHD do not result in failure on the WMT effort subtests in children of 7 years of age or more, as long as they make an effort and as long as reading ability is at grade level 3 or above, as measured by the Test of Academic Performance (Adams et al., 1989).

If we exclude the 5 children who initially failed the WMT but who passed it when given an incentive, 130 children remain. In these children, all with significant clinical problems, scores at the fifth percentile for the sample would be 82.5% for WMT IR, 82.5% for WMT DR and 75% for consistency. The latter WMT scores may serve as preliminary reference points when interpreting the scores from other children and they also provide comparative data, which are helpful in interpreting WMT effort subtest scores from adults. For example, we would probably doubt the validity of test results from an adult with a mild head injury or depression, if the person's WMT effort scores were lower than the fifth percentile relative to the children in the current study. The values of WMT IR and DR, which fall at the fifth percentile in these children are precisely the recommended adult cut-offs, below which effort should be considered inadequate to produce valid test data. The adult cut-offs were set at the second percentile relative to 157 patients with moderate to severe brain injury or neurological diseases affecting the brain, after excluding 10% of cases in whom there was strong independent evidence of poor effort (e.g., worse than chance scores on other tests).

The current findings need to be independently replicated. However, based on these preliminary findings, the WMT promises to be able to provide enhanced methodological sophistication in clinical testing and in research studies with children, as well as adults. It allows the simultaneous measurement of children's verbal memory and the effort that they are applying to testing. In children with a reading level of grade 3 or higher and aged 7 years or above, WMT effort scores at or below the fifth percentile relative to the current clinical sample may throw doubt on the effort applied to testing. The current findings may assist the clinician in interpreting WMT data from children of given diagnoses, ages, verbal intelligence levels and reading levels. In group studies, those with WMT scores below the fifth percentiles relative to the current clinical sample may be analyzed separately as cases demonstrating probable poor effort, assuming a reading grade level of 3 or above. Whereas the failure rate on the WMT in this sample was only 14%, which is much lower than typically found in groups of adults seeking compensation (Larrabee, 2000; Green, Rohling et al., 2001), even a 14% incidence of inadequate effort could distort the results of studies based on group comparisons and lead to erroneous conclusions.

## ACKNOWLEDGEMENTS

Thanks for critical feedback go to Drs. David Hartman and Roger Gervais.

## REFERENCES

Adams, W., Sheslow, D., & Erb, L. (1989). *The test of academic performance: Examiner's manual*. San Antonio: The Psychological Corporation.

Allen, L., & Green, P. (1999). CARB and WMT Effort test scores in 57 patients with severe traumatic brain injury. *Archives of Clinical Neuropsychology, 14*, 789.

Barkley, R.A. (1999). Theories of attention-deficit/hyperactivity disorder. In H.C. Quay & A.E. Hogan (Eds.), *Handbook of disruptive behavior disorders* (pp. 295–313). New York: Kluwer Academic Publishers Press/Plenum.

Bigler, E. (1990). Neuropsychology and Malingering: Comment on Faust, Hart, and Guilmette (1988). *Journal of Consulting and Clinical Psychology, 58*, 244–247.

Boa, B., Pelham, W.E., Waschbusch, D.A., Kipp, H., & Owens, J. (2001). Academic task persistence of normally achieving ADHD and control boys' performance; self-evaluations and attributions. *Journal of Consulting and Clinical Psychology, 69*, 271–283.

Dikmen, S.S., Machamer, J.E., Winn, H.R., & Temkin, N.R. (1995). Neuropsychological outcome at one year post head injury. *Neuropsychology, 9*, 80–90.

Douglas, V.I. (1983). Attentional and cognitive problems. In M. Rutter (Ed.), *Developmental neuropsychiatry* (pp. 49–66). New York: Guilford Press.

Faust, D., Hart, K., Guilmette, T.J., & Arkes, H.R. (1988a). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*, 508–515.

Faust, D., Hart, K., Guilmette, T.J., & Arkes, H.R. (1988b). Pediatric Malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *56*, 578–582.

Flaro, L., & Green, P. (2000). Symptom validity test results with children: CARB and WMT. *Archives of Clinical Neuropsychology*, *15*, 840.

Gervais, R., Green, P., Allen, L.M., & Iverson, G. (2001). Effects of coaching on symptom validity testing in chronic pain patients presenting for disability assessments. *Journal of Forensic Neuropsychology*, *2*, 1–19.

Gervais, R.O., Russell, A.S., Green, P., Allen, L.M., Ferrari, R., & Pieschl, S.D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology*, *28*, 1892–1899.

Green, P., Allen, L., & Astner, K. (1996). *Manual for Computerised Word Memory Test, U.S. version 1.0*. Durham, NC: Cognisyst.

Green, P., & Allen, L.M. (1999). Performance of Neurological Patients on the Word Memory Test (WMT) and Computerized Assessment of Response Bias (CARB): *Supplement to the Word Memory Test and CARB '97 Manuals*. Durham, NC: Cognisyst.

Green, P., & Astner, K. (1995). *The Word Memory Test*. Edmonton: Neurobehavioural Associates.

Green, P., Iverson, G., & Allen, L. (1999). Detecting Malingering in Head Injury Litigation with the Word Memory Test. *Brain Injury*, *13*, 813–819.

Green, P., Lees-Haley, P., & Allen, L. (2002). The Word Memory Test and the Validity of Neuropsychological Test Scores. *Forensic Neuropsychology*, *2*, 97–124.

Green, P., Rohling, M.L., Lees-Haley, P.R., & Allen, L.M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*, 1045–1060.

Hart, K.J. (1995). The assessment of malingering in neuropsychological evaluations: Research-based concepts and methods for consultants. *Consulting Psychology Journal: Practice and Research*, *47*, 246–254.

Hartman, D.E. (2002). The unexamined lie is a lie worth fibbing: Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology*, *17*, 709–714.

Hoza, B., Pelham, W.E., Waschbusch, D.A., Kipp, H., & Owens, J. (2001). Academic task persistence of normally achieving ADHD and control boys performance, self-evaluations, and attributions. *Journal of Consulting and Clinical Psychology*, *69*, 271–283.

Iverson, G., Green, P., & Gervais, R. (1999). Using the Word Memory Test to detect biased responding in head injury litigation. *Journal of Cognitive Rehabilitation*, *17*, 4–8.

Larrabee, G.J. (2000). Neuropsychology in personal injury litigation. *Journal of Clinical and Experimental Neuropsychology*, *22*, 702–707.

McCaffrey, R.J., & Lynch, J.K. (1992). A methodological review of "Method Skeptic" reports. *Neuropsychology Review*, *3*, 193–214.

McCann, J.T. (1998). *Malingering and deception in adolescents: Assessing credibility in clinical and forensic settings*. Washington, DC: American Psychological Association.

Meyers, J.E., & Meyers, K.R. (1995). *Rey Complex Figure Test and Recognition Trial manual*. Florida: Psychological Assessment Resources, Inc.

Meyers, J.E., & Meyers, K.R. (1996). *Rey Complex Figure Test and Recognition Trial: Supplemental norms for children and adolescents*. Florida: Psychological Assessment Resources, Inc.

Rogers, R. (1997). Current Status of Clinical Methods. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 373–397). New York: Guildford Press.

Schmitz, B., & Skinner, E. (1993). Perceived control, effort, and academic performance: Interindividual, intraindividual, and multivariate time series analysis. *Journal of Personality and Social Psychology*, *64*, 1010–1028.

Talley, J.L. (1993). *Children's Auditory Verbal Learning Test* (2nd ed.). Florida: Psychological Assessment Resources, Inc.