Two studies examined the Warrington Recognition Memory Test (RMT) discrepancy index (Words-Faces) in a large sample of patients heterogeneous with respect to age, education, gender, and neurological diagnosis. In Study 1 (N = 504) we used cutoffs from the Words-Faces discrepancy scores derived from Warrington’s original validation sample to attempt to accurately classify patients with left, right, or diffuse brain damage. Sensitivity for left hemisphere patients (Faces > Words) was 10% with a specificity of 88%, whereas sensitivity for right hemisphere patients (Words > Faces) was 48% with a specificity of 86%. For patients with diffuse brain damage (Words = Faces) sensitivity was 69% and specificity was 38%. In Study 2 (N = 263), we examined the relationship between the Words-Faces discrepancy score and Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1981) Logical Memory and Visual Reproduction subtests. Contrary to predictions, patients with Words > Faces performed better on both WMS-R subtests; the Faces > Words discrepancy was not related to Visual Reproduction performance. Potential reasons for these negative findings are discussed, as well as cautions for future RMT discrepancy index use. © 2000 National Academy of Neuropsychology. Published by Elsevier Science Ltd

Keywords: recognition memory, Warrington, material specificity

The Warrington Recognition Memory Test (RMT; Warrington, 1984) is a widely used measure of recognition memory for verbal (Words) and nonverbal (Faces) material. In
addition to separate measures of Words and Faces performance, a discrepancy score (Words-Faces) can be computed to determine an individual’s relative memory strength or weakness. According to Warrington (1984), left hemisphere-damaged patients are expected to perform significantly better on Faces compared to Words, and right hemisphere patients better on Words than Faces.

Despite its frequent use, a number of psychometric concerns have been raised about the RMT and a small body of research with epileptic patients does not support its purported ability to be sensitive to lateralized impairment. For instance, in a sample of patients with temporal lobe epilepsy, Hermann, Connell, Barr, and Wyler (1995) found that only 15% of left-hemisphere patients and 10% of right-hemisphere patients were correctly classified based on the Words-Faces discrepancy score. These rates improved to 42% and 31%, respectively, 6 to 8 months following anterior temporal lobectomy. Specificity, or true negatives, was uniformly high at about 90%. Roughly similar classification rates were obtained by Naugle, Chelune, Schuster, Luders, and Comair (1994) pre- and postanterior temporal lobectomy, but Kneebone, Chelune, and Luders (1997) obtained lower posttemporal lobectomy classification rates of 8% for left-hemisphere patients and 15.6% for right-hemisphere patients. In addition to this epilepsy research, careful examination of Warrington’s (1984) large validation sample of vascular and tumor patients also indicates limited sensitivity. Only 29% of right hemisphere patients and 28% of left hemisphere patients demonstrated the expected Words versus Faces discrepancy. In summary, although some research demonstrates that the RMT may be sensitive to Alzheimer’s disease (Diesfeldt, 1990), it appears relatively insensitive to lateralized brain impairment, at least among epilepsy samples.

In addition to this limited classification data, Kapur (1987) raised a number of psychometric concerns about the RMT. For instance, no reliability data of any type is presented in the manual, the test does not have parallel forms, and there is a significant ceiling effect on the Words subtest. Because of this ceiling, Naugle et al. (1994) argued that comparison of the Words-Faces discrepancy index across pre- and posttesting potentially can be problematic. A recent study by Malina, Bowers, Millis and Uekert (1998) demonstrated that the RMT subtests have adequate internal consistency, with Cronbach’s alpha of .86 for Words and .77 for Faces, respectively.

In addition to psychometric concerns, limited research has examined the relationship of the RMT with other supposedly material-specific measures, such as versions of the Wechsler Memory Scale (WMS; Wechsler, 1945). One study by Compton, Sherer, and Adams (1992) found, contrary to what might be expected given the material similarity, that the RMT Words and Faces subtests did not load on the verbal or figural factors, respectively, with other WMS subtests. Rather, both primarily loaded on a separate recognition memory factor.

Overall, review of studies to date indicates that basic psychometric properties and concurrent validity of the RMT have not been demonstrated adequately and deserve further scientific study. Also, although a considerable amount of research has assessed the RMT in epileptics, relatively little research has examined performance in patients with lateralized disorders of other etiology or patients with diffuse brain impairment.

Clinically, the present investigators had been using the RMT at various sites (to be described below). After repeatedly observing instances of unpredicted and apparently nonsensical performance patterns in the clinical setting (e.g., intact free recall on multiple tasks, with impaired RMT; significant Words-Faces discrepancies in patients with diffuse brain damage), systematic examination of the RMT appeared important. The current investigation was conducted to assess RMT performance in a heterogeneous sample of patients of varying ages, educational levels, races, and neurological diagnoses, with two
broad goals. The first goal was to evaluate the sensitivity of the RMT discrepancy score to the presence of lateralized lesions across a broad sample of patients. Based on Warrington (1984), it was expected that left hemisphere-damaged patients would perform significantly worse on Words versus Faces and that right hemisphere patients would perform significantly worse on Faces versus Words. The second goal was to determine the relationship of the RMT discrepancy score to other presumably material specific tests, as previous research has demonstrated a relationship between verbal and nonverbal WMS-R stimuli and laterality of brain dysfunction (see Jones-Gotman, 1991). It was expected that patients who performed significantly better on Words versus Faces would also perform better on WMS-R Logical Memory, whereas patients who performed significantly better on Faces versus Words would perform better on WMS-R Visual Reproduction. Goal 1 was addressed in Study 1 and goal 2 was addressed in Study 2.

STUDY 1

Method

Patients. Patients ($N = 504$) were retrospectively obtained from three sites—the Neuropsychology Service of Evanston Hospital (EH: Evanston, IL) and the Traumatic Brain Injury and Stroke Programs of the Rehabilitation Institute of Michigan (RIM: Detroit, MI). All patients were consecutive referrals for neuropsychological evaluation to the three respective neuropsychology services and had independent confirmation of brain damage, typically confirmed by computerized tomography (CT) scan, magnetic resonance imaging (MRI) or electroencephalogram (EEG). The EH sample ($n = 251$) consisted of the following patients; 17 tumor, 50 stroke, 61 closed-head injury, 7 seizure disorder, 10 dementia, 1 multiple sclerosis, 26 Parkinson’s disease, 40 other/multiple diagnoses, and 39 etiology unknown. RIM samples were divided into three groups by etiology. RIM1 consisted of 110 stroke patients (51 left and 59 right), RIM2 consisted of 60 accident victims (34 assault, 16 fall, and 10 pedestrians hit by vehicles), and RIM3 consisted of 83 patients who sustained a closed-head injury in motor vehicle accidents. RIM2 and RIM3 patients were obtained from the same clinical service, but differed in type of brain injury. Basic demographic information on these patients is presented in Table 1.

A one-way analysis of variance (ANOVA) indicated a significant age difference between groups, $F(3, 500) = 134.1, p < .0001$. Tukey’s HSD post-hoc test indicated significant age differences between all groups ($p < .05$) in the following order (from oldest to

<table>
<thead>
<tr>
<th>Sample</th>
<th>$n$</th>
<th>Age $M$ (SD)</th>
<th>Education $M$ (SD)</th>
<th>White (%)</th>
<th>Male (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EH</td>
<td>251</td>
<td>54.4 (17.8)</td>
<td>14.7 (2.7)</td>
<td>95</td>
<td>56</td>
</tr>
<tr>
<td>RIM1</td>
<td>110</td>
<td>67.4 (6.6)</td>
<td>10.7 (1.6)</td>
<td>41</td>
<td>51</td>
</tr>
<tr>
<td>RIM2</td>
<td>60</td>
<td>37.5 (13.6)</td>
<td>11.6 (2.2)</td>
<td>38</td>
<td>85</td>
</tr>
<tr>
<td>RIM3</td>
<td>83</td>
<td>28.7 (11.3)</td>
<td>12.0 (1.8)</td>
<td>54</td>
<td>70</td>
</tr>
</tbody>
</table>

EH = mixed etiology; RIM1 = cerebrovascular accidents from the Stroke Program at the Rehabilitation Institute of Michigan; RIM2 = Non-motor vehicle accident (MVA) from the Brain Injury Program at the Rehabilitation Institute of Michigan; RIM3 = MVA closed-head injury from the Brain Injury Program of the Rehabilitation Institute of Michigan. See text for complete description of patients.
youngest); RIM1 patients, EH patients, RIM2 patients, and RIM3 patients. These age differences are a function of etiology, as the RIM1 sample consisted of stroke patients, RIM2 and RIM3 were patients with traumatic brain injury, and EH patients were heterogeneous with respect to etiology. Another one-way ANOVA indicated a significant education difference between groups, $F(3, 497) = 159.7, p < .00001$, with EH patients significantly more educated ($p < .05$) than the other groups and RIM1 patients less educated than RIM3 patients ($p < .05$). EH patients were also more likely to be Caucasian than the other groups $\chi^2 (3, n = 497) = 157.16, p < .000001$; RIM2 patients were more likely to be male than the other groups $\chi^2 (3, n = 503) = 24.9, p < .00002$. This latter finding is consistent with the known risk factor of male gender in traumatic brain injury. These demographic analyses indicate that the final sample consisted of heterogeneous neurological patients with respect to age, education, and gender.

In addition to these differences between samples, we also analyzed age and education differences between the laterality groups (see Table 2). Two separate one-way ANOVAs indicated that the groups did not differ based on age, $F(2, 501) = .61, ns$, but differed on educational attainment, $F(2, 498) = 44.98, p < .00001$. Tukey’s HSD post-hoc test indicated that the diffusely damaged group was significantly more educated than the right or left hemisphere groups, who did not differ from one another. This finding is due to the fact that most of these patients (207/259) were from the EH sample that, as noted above, was more educated that RIM samples.

Procedure. All patients were administered the RMT according to standardized instructions (Warrington, 1984) as part of a neuropsychological evaluation.

Results. A mixed-model analysis of covariance (ANCOVA) was completed on patients from all samples with laterality (left, right, and diffuse) as the between-subjects factor and RMT subtest (Words and Faces) as the within-subjects factor. Because of the education differences between groups, we used education as a covariate. Means and standard deviations of the RMT variables, as well as demographic characteristics of the three groups, are presented in Table 2. The main effect of laterality was not significant, $F(2, 497) = 1.85 ns$, indicating that there were no significant differences between groups across RMT subtests, but the main effect of RMT subtest was significant, $F(1, 498) = 193.59, p < .0001$, with a significant advantage for Words compared to Faces. The inter-
action effect was also significant, $F(2, 498) = 42.31, p < .0001$. Tukey’s post-hoc tests ($p < .05$) indicated that the right hemisphere patients performed better on Words relative to the left hemisphere and diffusely damaged patients, whereas the left hemisphere group performed significantly better on Faces than the right hemisphere group, but similarly to the diffusely damaged group. We conducted similar ANOVA analyses for each sample separately and obtained the same results for the main effects of laterality and RMT subtest. Similar results at each RIM sample, but not the EH sample, were also obtained for the laterality $\times$ RMT subtest interaction; at each site, right hemisphere patients performed better than left hemisphere patients on the Words subtest, but only in the RIM1 sample (stroke patients) did left hemisphere patients perform better on Faces than the right hemisphere patients. These findings partially support Warrington’s contention that patients with lateralized damage tend to have selective memory deficits on RMT subtests, particularly for left hemisphere patients on the Words subtest.

Although analysis of group means has been used frequently to assess patient performance on various neuropsychological tests, more recent attention has focused on effectiveness in classifying neuropsychological or neurological disorders (Barr, 1997; Guilmette & Rasile, 1995; Monsch, Bondi, Butters, Salmon, Katzman & Thal, 1992). As such, we conducted classification analyses for the Words-Faces discrepancy score. This score was obtained by subtracting Faces from Words and comparing the score to normative data presented in the RMT manual. Per Warrington’s (1984) recommendations, a Words-Faces discrepancy score was considered to be significant (i.e., a selective deficit) if it was present in less than 5% of the standardization sample. For Words to be significantly greater than Faces, 10 or more words than faces were required; for Faces to be significantly greater than Words, 6 or more faces than words were required. In this manner, we compared actual lateralized damage to the discrepancy score to obtain diagnostic efficiency statistics (presented in Table 3). It should be noted that these efficiency statistics are somewhat different than traditional diagnostic measures in that they have been applied to three distinct neurologically impaired groups, rather than to the pre-

### TABLE 3
Diagnostic Efficiency Statistics for the Warrington Recognition Memory Test (RMT) Faces $>$ Words for Identifying LH Patients, RMT Words $>$ Faces for Identifying RH Patients, and RMT Words $=$ Faces for Identifying Diffusely Damaged Patients

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Area of Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left Hemisphere$^a$</td>
</tr>
<tr>
<td>Overall hit rate</td>
<td>.70</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>.10</td>
</tr>
<tr>
<td>Specificity</td>
<td>.88</td>
</tr>
<tr>
<td>Positive predictive power</td>
<td>.20</td>
</tr>
<tr>
<td>Negative predictive power</td>
<td>.77</td>
</tr>
<tr>
<td>False-positive rate</td>
<td>.12</td>
</tr>
<tr>
<td>False-negative rate</td>
<td>.90</td>
</tr>
</tbody>
</table>

---

Hit rate = proportion of patients correctly identified (true positive + true negative/N); Sensitivity = true positive rate, proportion of patients with target laterality correctly identified; Specificity = true negative rate, proportion of patients not having the target laterality correctly identified as not having that laterality; positive predictive power = probability that the patient has the target laterality given that the test identifies patient as such; Negative predictive power = probability that the patient does not have the target laterality given that the test does not identify patient as such.

$^a$RMT Faces $>$ Words used to identify left hemisphere patients; $n = 115$.

$^b$RMT Words $>$ Faces used to identify right hemisphere patients; $n = 130$.

$^c$RMT Words $=$ Faces used to identify diffusely damaged patients; $n = 259$. 

Downloaded from https://academic.oup.com/acn/article-abstract/15/4/301/1900 by guest on 10 March 2019
ence or absence of a disorder or disease, as is most frequently done. As can be seen in Table 3, the sensitivity of the Words-Faces discrepancies varies, from a low of 10% for left hemisphere patients when Words < Faces, to a high of 69% for patients with diffuse brain injury when Words = Faces. Sensitivity for right hemisphere patients was 48% when Words > Faces. Specificity, or true negatives, was quite low for the diffusely injured patients (38%), but substantially higher for the left and right hemisphere damaged patients (88% and 86%, respectively). Further analyses were conducted on the patients with diffuse brain injury because a Words = Faces result is not necessarily an indication of impairment, in contrast to a significant Words versus Faces discrepancy. Of the 259 patients with diffuse brain injury, 55 scored below cutoffs indicative of impairment (i.e., the 5th percentile of the standardization sample) on both Words and Faces subtests. We separately computed diagnostic efficiency statistics on these patients and found a sensitivity of 73% and an overall hit rate of 44%. Specificity, because it is computed on patients with only lateralized damage, remained at 38%. Because these classification rates are similar to rates for the entire diffusely injured sample, we did not further compute diagnostic efficiency statistics on this subsample of patients.

Two further analyses were conducted. First, when patients were separated by sample, we found roughly similar overall levels of classification accuracy for patients with left, right, and diffuse brain damage compared to the entire sample, although ranges of sensitivity and specificity varied. For left hemisphere patients, sensitivity ranged from 4 to 14% and specificity ranged from 84 to 100%. For right hemisphere patients, sensitivity ranged from 33 to 53% and specificity ranged from 83 to 96%. For patients with diffuse brain injury, sensitivity ranged from 68 to 77% and specificity ranged from 38 to 48%. Second, we also computed diagnostic efficiency statistics for the 10th percentile Words-Faces discrepancy indicated by Warrington (1984). For left hemisphere patients, sensitivity was 19%, specificity was 87%, and the overall hit rate was 72%, whereas for right hemisphere patients sensitivity was 71%, specificity was 71%, and the overall hit rate was 71%. For patients with diffuse damage, sensitivity was 51%, specificity was 58%, and the overall hit rate was 55%. Compared to the 5% discrepancy score, sensitivity increased for left and right hemisphere damaged patients at the expense of specificity. For instance, for patients with right hemisphere damage, sensitivity increased from 48 to 72%, while specificity dropped from 86 to 71%. The reverse pattern of change was evident with diffusely damaged patients, for whom sensitivity decreased from 69 to 58% (as the range of Words-Faces scores considered significantly discrepant decreased), while specificity increased from 38 to 58%. The overall hit rate remaining similar (54–55%) for the diffusely damaged group.

Compared to the base rate, the obtained sensitivity of 10% for left hemisphere damage using the 5% discrepancy cutoff in the whole sample is actually below expectations (115/504 = 23%). However, the obtained sensitivity of 48% for right hemisphere damage is above base rate expectations (130/504 = 26%) and the obtained sensitivity of 69% is above the base rate of patients with diffuse brain damage (273/504 = 54%). In all, these findings indicate that the Words-Faces discrepancy is generally insensitive to lateralized neurological impairment (particularly for left hemisphere damage), but somewhat better in the detection of diffuse brain dysfunction when no RMT discrepancy is found.

**STUDY 2**

Methods

Patients. Only EH patients (n = 251) who had been administered both the RMT and the WMS-R were used in this study. This test data was only available from this site.
Patients were administered the RMT and WMS-R subtests of Logical Memory and Visual Reproduction according to standardized instructions in the context of a neuropsychological evaluation. WMS-R scores were converted into z-scores, based on norms provided in the manual. If the participant did not fit into WMS-R age categories, the closest category was used, except for patients aged 25 to 34 years, for whom norms provided in Mittenberg, Burton, Darrow, and Thompson (1992) were used. Three separate participant groups were created based on RMT discrepancy scores according to information contained in the manual; Words > Faces, Faces > Words, and Words = Faces.

Results. Basic demographic information on the three groups, as well as the means and standard deviations (expressed in z units) for Logical Memory I and II and Visual Reproduction I and II are presented in Table 4. A one-way ANOVA indicated a significant age difference between groups $F(2, 248) = 8.8, p < .0002$. The Faces > Words patients were significantly older than both other groups, but the other groups did not differ from one another. Educational attainment and gender composition was similar for all groups. We next conducted a mixed model ANCOVA with one between subjects factor (RMT grouping), two within-subjects factors (Logical Memory and Visual Reproduction), and age as the covariate. Each of the within subjects factors had two levels (i.e., immediate and delayed recall). The main effect of RMT grouping was significant, $F(2, 243) = 4.34, p < .01$, with the Words > Faces groups performing significantly better than the other groups. The within-subjects main effect of Logical Memory was not significant, but the main effect of Visual Reproduction was significant, $F(1, 244) = 43.22, p < .0001$. This latter finding indicates that patients performed significantly better on immediate versus delayed visual recall. Respective interaction effects between RMT grouping and both Logical Memory and Visual Reproduction were not significant and the three-way interaction between RMT grouping, Logical Memory, and Visual Reproduction was not significant. Overall, these findings indicate that RMT grouping is not related in the expected fashion to material specific subtests from the WMS-R.

DISCUSSION

In these studies, we have demonstrated that the Words-Faces discrepancy has poor diagnostic efficiency for laterality of brain damage and is not related in the expected fashion to WMS-R subtests. Although considerable research has assessed the Words-
Faces discrepancy in epileptic populations, this study is the first to assess this discrepancy in a large group of patients heterogeneous with respect to age, education, gender, and neurological diagnosis. The broad scope of our inquiry is therefore likely to be useful to the clinician who typically evaluates many different types of patients in whom the diagnosis is not known.

Study 1 demonstrates an increasingly common method of analyzing data with use of diagnostic efficiency statistics rather than group means (see Retzlaff & Gibertini, 1994). If one were to look simply at group means, there is some evidence that the RMT may be sensitive to lateralized damage, as right hemisphere patients performed significantly better on Words versus Faces. Yet, when diagnostic efficiency statistics are applied to the discrepancy data, the classification rates are quite low and, in some case, below base rate expectations. Sensitivity for right and left hemisphere damage (10% and 48%, respectively) are too low to use confidently in a clinical setting. Sensitivity for diffusely brain damaged patients is higher (69%), but it is also accompanied by a high false-positive rate (62%). These data suggest that the RMT’s discrepancy index of a 5 percentile cutoff suggested by Warrington (1984), as well as the more liberal 10 percentile cutoff, are not useful in detecting lateralization of brain damage. Barr (1997) recently came to a similar conclusion about the WMS-R’s ability to correctly classify patients with lateralized temporal lobe seizure onset.

Another of our main findings was that the RMT discrepancy did not show the expected pattern of relationships with WMS-R subtests of similar material specificity. Similarly, Compton et al. (1992) performed a factor analysis with the RMT and WMS on a large sample of clinical referrals and found both RMT subtests to load on a separate recognition memory factor. Neither loaded on verbal or figural memory factors with WMS subtests. These negative findings, in conjunction with our own, raise an important question of what the RMT is actually measuring, if it is not sensitive to lateralized damage and not related to material specific memory functioning. Although we do not yet have the data to address this issue, if the RMT is to remain a frequently used neuropsychological test, future research is needed to clarify and elaborate its relationship with other measures. Factor-analytic studies, with both large clinical samples and nonneurologically impaired, may be particularly useful in this regard.

One of the main limitations of this study is that a global measure of hemispheric damage, either right or left, was used rather than the specific lobe. One might predict that temporal lobe patients would demonstrate the most pronounced memory impairment, compared to other areas of damage. Warrington’s (1984) breakdown of left and right hemisphere patients did, in fact, find this trend but only 39.6% of the right temporal patients showed the expected Words > Faces discrepancy, and only 35.4% of the left temporal patients showed the expected Faces > Words discrepancy. Although somewhat better than classification rates for the other damaged lobes, these rates are quite low for clinical use. As a final caveat, the rather artificial nature of these studies, in which the performance of only one test in isolation is assessed, should be considered. Rarely is the RMT (or any other neuropsychological test) administered alone to determine site of brain dysfunction. Rather, the measure is typically used in conjunction with several other neuropsychological measures, as well as other data points, including behavioral signs, clinical interview, etc. As such, although our data clearly indicate that the RMT is not sensitive enough to use in isolation to determine lesion laterality, we cannot rule out that the RMT may be useful when combined with other neuropsychological measures in a multifactorial methodology to determine lesion laterality. Also, we note that some research has indicated that it can be useful in the detection of malingering or insufficient effort (Millis, 1992).
REFERENCES


