

The Homophone Meaning Generation Test: Psychometric properties and a method for estimating premorbid performance

J.R. CRAWFORD¹ AND ELIZABETH K. WARRINGTON²

¹Department of Psychology, University of Aberdeen, Aberdeen, Scotland, UK

²Dementia Research Group, Institute of Neurology, London, UK

(RECEIVED December 20, 2000; REVISED May 17, 2001; ACCEPTED May 24, 2001)

Abstract

The Homophone Meaning Generation Test (HMGT; Warrington, 2000) is a new measure of verbal fluency that has been demonstrated to be sensitive to the presence of anterior lesions. In the present study we used the HMGT healthy standardization sample ($N = 170$) and demonstrate that scores on the HMGT do not differ significantly from a normal distribution and that the test has adequate reliability ($\alpha = .82$). A table for obtaining confidence limits on an individual's score is presented. A regression equation for the estimation of premorbid HMGT performance was constructed using the National Adult Reading Test as the predictor variable. In a sample of 36 cases with anterior lesions estimated premorbid scores were significantly higher than obtained scores ($p < .001$). Premorbid ability acted to suppress group differences on the HMGT; the partial correlation between neurological status (healthy vs. anterior lesion) and HMGT performance controlling for premorbid ability (.53) was significantly higher than the raw correlation (.44). In addition, hierarchical discriminant function analysis demonstrated that the inclusion of premorbid ability improved classification over that achieved by HMGT scores alone. These results support both the underlying rationale and the clinical utility of controlling for premorbid performance when interpreting verbal fluency scores. (*JINS*, 2002, 8, 547–554.)

Keywords: Verbal fluency, Premorbid ability, Executive function, Frontal lobes

INTRODUCTION

Executive dysfunction is considered to underlie many of the behavioral changes observed in a wide range of neurological and psychiatric disorders. However, although clinicians regularly observe its often devastating effects on the capacity for independent living, it has proved very difficult to develop reliable and valid methods of quantifying such dysfunction.

Measures of verbal fluency have commonly been used to assess executive dysfunction (Crawford et al., 1998; McCarthy & Warrington, 1990). Warrington (2000) has recently designed a new measure of verbal fluency termed the Homophone Meaning Generation Test (HMGT). This task requires participants to generate multiple meanings for

each of a series of eight homophones (e.g., *sent*, *tick*, etc.). It is argued that this task requires greater cognitive flexibility than existing fluency measures as it requires multiple switches between verbal concepts.

Warrington (2000) demonstrated that performance on this task is impaired following anterior lesions, regardless of laterality. Other tasks, most obviously card sorting tests such as the Wisconsin Card Sorting Test (Grant & Berg, 1948; Heaton, 1981) and Modified Card Sorting Test (Nelson, 1976), also require shifting of cognitive set. However, the number of shifts in these former tasks are limited. In addition, scores on these tasks are heavily skewed, to the extent that they could be regarded as simply providing a pass or fail measure. Warrington (2000) argues that tests of executive dysfunction which yield normally distributed scores have many advantages and she asserts that the HMGT should possess this property. Burgess and Shallice (1997) have also presented a convincing case for developing normally distributed tests of executive functioning.

Reprint requests to: Professor John R. Crawford, Department of Psychology, King's College, University of Aberdeen, Aberdeen, Scotland, AB29 2UB, UK. E-mail: j.crawford@abdn.ac.uk

The present study evaluates the measurement characteristics of the HMGT and provides data to aid clinicians with interpretation of an individual's performance. The first specific aim is to test Warrington's (2000) assertion that scores on the HMGT will be normally distributed in the healthy population. The second issue examined is the reliability of the test. Adequate reliability is a necessary condition for validity and is particularly crucial when, as in the case of the HMGT, a test is intended for use in the individual case (Crawford, 1996; Franzen, 1989). We use Cronbach's alpha to estimate reliability of the HMGT and subsequently to generate confidence limits on individuals' HMGT scores. Confidence limits serve the general purpose of reminding users that test scores are fallible but they also quantify the *degree* of uncertainty; because of this their use is strongly recommended by a number of authorities (e.g., Nunnally & Bernstein, 1994; Stanley, 1971).

Performance on verbal fluency tests is strongly related to verbal IQ in the general population. For example, Crawford et al. (1993) reported a correlation of .64 between initial-letter fluency and Wechsler Verbal IQ in a healthy sample ($N = 144$). Some studies have reported even higher correlations; for example, Miller (1984) reported a correlation of .86 between fluency and Verbal IQ in a small, healthy sample. Such results indicate that an individual's premorbid ability should be considered when interpreting verbal fluency performance in clinical populations. This was graphically illustrated by Borkowski et al. (1967) who reported that the fluency performance of a brain-damaged sample of above-average Verbal IQ was significantly higher than that of healthy subjects of below-average Verbal IQ.

Crawford et al. (1992) employed a healthy sample to build a regression equation for the estimation of premorbid initial-letter verbal fluency performance from scores on the National Adult Reading Test (NART; Nelson & Willison, 1991). They suggested that estimated fluency performance should be compared to obtained fluency; a large discrepancy between the two in favor of premorbid ability would constitute evidence of an acquired fluency deficit.

The NART and its variants are widely used to estimate premorbid ability. Although NART performance is impaired in severe dementia and in some other clinical disorders, in general, performance is surprisingly robust in the face of neurological and psychiatric illness (see Crawford, 1992; Franzen et al., 1997; and O'Carroll, 1995, for reviews).

Crawford et al. (1992) reported that the NART had a highly significant correlation ($r = .67$) with initial-letter fluency and provided a table to convert NART errors to estimated fluency performance. They also provided provisional evidence of validity for the method by demonstrating a highly significant difference between estimated premorbid fluency and obtained fluency in a neurological sample.

Warrington (2000) reported that the NART had a highly significant correlation with HMGT performance ($r = .60$). This suggests that, as in the case of initial-letter fluency, a NART equation for the estimation of premorbid performance would be a useful supplement to conventional HMGT

norms as it will provide an *individual* comparison standard (Lezak, 1995) against which to compare a patient's obtained HMGT scores. In the present study we use the healthy standardization sample for the HMGT to generate this equation and provide additional data to permit its use in clinical practice.

We also evaluate the validity and utility of the equation in three ways. Firstly, we test the hypothesis that obtained HMGT scores will be significantly lower than estimated premorbid HMGT scores in a sample of cases with anterior cortical lesions; this is directly analogous to Crawford et al.'s. (1992) evaluation of the equation for estimating premorbid performance on initial-letter fluency. Secondly, we use hierarchical discriminant function analysis to test whether incorporating estimates of premorbid performance improves discrimination between healthy and anterior lesion cases over that achieved by the HMGT alone. Finally, we examine whether premorbid ability acts to suppress the relationship between fluency performance and neurological status. The hypothesis tested is that the raw (point-biserial) correlation between fluency and neurological status (i.e., healthy vs. anterior lesion) will be significantly lower than the partial correlation obtained after controlling for premorbid ability.

METHODS

Research Participants

Two samples were employed. The first sample consisted of the 170 healthy participants (102 females, 68 males) recruited by Warrington (2000) to serve as the HMGT standardization sample. The mean age in this sample was 45.1 ($SD = 14.9$) with a range from 19 to 74 years. This sample was broadly representative of the adult UK population in terms of the distributions of age and socio-economic status; for further details see Warrington (2000).

The second sample consisted of 35 patients (21 males, 14 females) with verified focal anterior lesions tested by Warrington (2000). The majority of these patients had space occupying tumors, the remainder had well localized vascular lesions; in 17 of the cases the lesion was in the left hemisphere and in the right hemisphere in the remainder. Mean age of the sample was 45.8 ($SD = 14.0$). For further details of this sample see Warrington (2000).

Tests and Materials

Participants in both samples had been administered the HMGT and the NART according to standard instructions. The healthy sample has also been administered the Graded Naming Test (McKenna & Warrington, 1983), and the anterior lesion sample had been administered the Modified Card Sorting Test (Nelson, 1976); these data are not used in the present investigation.

The HMGT consists of eight homophones (*tick, tip, slip, form, plain, bored, right, sent*). It can be seen that some of the homophones have a single spelling (e.g., *slip*) while others have multiple spellings (e.g., *sent-scent-cent*). The homophones are presented orally and there is no time constraint. A point is awarded for each distinct meaning. Summary statistics for the individual HMGT items are presented in Table 1. The individual HMGT items are summed to obtain a raw score and this raw score can then be converted to a scaled score (i.e., $M = 10, SD = 3$).

The NART is an oral, single word reading test consisting of 50 words that violate grapheme-phoneme correspondence rules (e.g., *chord*). By convention, performance is expressed as the number of errors of pronunciation, with high scores therefore reflecting poor performance.

Analysis

To obtain 95% confidence limits on HMGT scaled scores the following formula was used to calculate the standard error of measurement for true scores (Glutting et al., 1987; Stanley, 1971):

$$SEM_{xt} = r_{xx}(S_x\sqrt{1 - r_{xx}}), \tag{1}$$

where S_x is the standard deviation of the scale (3 in the present case as we are working with HMGT scaled scores), and r_{xx} is the reliability of the scale (normally estimated using Cronbach’s alpha). Confidence limits are formed by multiplying the SEM by a value of z (a standard normal deviate) corresponding to the desired confidence limits; for 95% limits, the most commonly used, this value is 1.96.

These confidence limits are not symmetrical around individuals’ *obtained* scores but around their estimated *true* scores (Nunnally & Bernstein, 1994; Silverstein, 1989; Stanley, 1971). The estimated true score is obtained by multiplying the obtained score, in deviation form, by the reliability of the test. It can be seen then that true scores are regressed towards the mean, the extent of this regression varying inversely with the reliability of the scale. The formula is as follows:

$$\text{Estimated true score} = r_{xx}(X - \bar{X}) + \bar{X}, \tag{2}$$

Table 1. Summary statistics for individual HMGT items

Statistic	HMGT item							
	Tick	Slip	Tip	Form	Plain	Bored	Right	Scent
<i>M</i>	2.74	2.96	3.13	2.96	3.34	2.85	3.19	2.53
<i>SD</i>	1.04	0.98	0.99	1.05	0.81	0.97	0.86	0.64
Minimum	0	1	1	1	1	1	1	1
Maximum	5	5	6	6	5	6	5	4

where X is the obtained score and \bar{X} is the mean for the scale (10 in the present case). Thus, for example, if an individual obtained a score of 5 on a scale that had a mean of 10 and a reliability of .8, the individual’s estimated true score would be 6.

RESULTS

Distribution of Test Scores and Reliability

A Kolmogorov-Smirnov test applied to the distribution of HMGT scores in the healthy sample revealed that scores on the HMGT did not deviate significantly from a normal distribution; $z = 1.25, p = .09$. The internal consistency of scores on the HMGT was examined using Cronbach’s Coefficient Alpha (α). Alpha was .82; the 95% confidence interval on this alpha, calculated using Feldt’s (1965) formula, was .78 to .86. Cronbach’s alpha was entered into formulae (1) and (2) to generate estimated true scores and 95% confidence limits for HMGT scaled scores. These limits are presented in Table 2.

Regression Equation for the Estimation of Premorbid HMGT Performance

In the standardization sample the mean of HMGT raw scores was 23.7 ($SD = 4.9$) and mean NART errors was 22.9 ($SD = 9.3$). The correlation between the NART and the HMGT was .605 ($p < .001$). HMGT raw scores were regressed on NART error scores to generate the following equation:

Table 2. Table for obtaining 95% confidence limits for true scores on the HMGT

Scaled score	Estimated true score	95% confidence limits on true scores	
		Lower limit	Upper limit
1	3	1	5
2	3	1	5
3	4	2	6
4	5	3	7
5	6	4	8
6	7	5	9
7	8	5	10
8	8	6	10
9	9	7	11
10	10	8	12
11	11	9	13
12	12	10	14
13	12	10	15
14	13	11	15
15	14	12	16
16	15	13	17
17	16	14	18
18	17	15	19
19	17	15	19
20	18	16	20

Estimated premorbid HMGT performance = $30.1 - (0.318 \times \text{NART errors})$.

The standard error of estimate for the equation was 3.91. For ease of use, Table 3 converts NART error scores to estimated premorbid HMGT scores. The standard error of estimate for the HMGT was multiplied by *z* values of 1.03, 1.64 and 2.32 to derive the critical values for the 15%, 5% and 1% levels of significance. These critical values are presented in Table 4. Because clinicians or researchers using this equation will wish to test a directional hypothesis (i.e., that obtained scores are lower than estimated premorbid scores), the critical values are one-tailed.

Validity and Utility of the Regression Equation

Mean NART errors in the anterior lesion sample was 21.8 (*SD* = 9.36). Estimated premorbid HMGT performance for each anterior lesion case was calculated from NART errors using the regression equation. Mean estimated *premorbid* HMGT performance was 24.07 (*SD* = 2.98). Mean *obtained* scores on the HMGT in the anterior lesion sample was 17.2 (*SD* = 5.40). A paired samples *t* test was used to compare obtained scores with estimated premorbid scores. This revealed a highly significant difference in favor of estimated premorbid scores ($t = 7.81, df = 34, p < .001$).

A hierarchical discriminant analysis was performed to test whether the combination of estimated premorbid ability and HMGT scores would improve discrimination between the healthy and anterior lesion samples over that achieved by HMGT scores alone. The overall classification accuracy (i.e., the percentage of cases correctly classified) was 74.8% for HMGT scores alone (75%

Table 4. Critical values for the discrepancy between obtained scores and estimated premorbid scores on the HMGT

	Significance level (one-tailed)			
	.15	.10	.05	.01
Discrepancy	4.1	5.0	6.4	9.1

of controls and 72% of anterior lesion cases were correctly classified). When premorbid ability as estimated by the NART was included, the overall classification accuracy rose to 80.5% (82% of controls and 74% of anterior lesion cases were correctly classified). A McNemar repeated measures chi-square test was used to test whether the improvement in classification accuracy was statistically significant (Tabachnick & Fidell, 1996). To perform this test, the number of cases initially correctly classified by HMGT scores alone *and* subsequently incorrectly classified with the addition of NART ($n = 9$), was compared to the number of cases for whom the converse occurred ($n = 21$). This test revealed a significant improvement in classification accuracy ($\chi^2 = 4.03, df = 1, p = .023$).

The correlations between premorbid ability (as measured by the NART), fluency (as measured by the HMGT) and neurological status are reported in Figure 1 along with their significance levels. In coding neurological status, healthy cases were assigned a value of zero and lesion cases a value of 1. To ease interpretation, NART scores were reflected for this part of the analysis so that high scores represented good performance.

It can be seen from Figure 1 that there is a significant (point-biserial) correlation between neurological status and the HMGT. This demonstrates a significant between-group difference in HMGT performance in favor of the controls

Table 3. Table for converting NART errors to estimated premorbid HMGT performance

NART errors	Premorbid HMGT	NART errors	Premorbid HMGT	NART errors	Premorbid HMGT
0	30.1	17	24.7	34	19.3
1	29.8	18	24.4	35	19.0
2	29.5	19	24.1	36	18.7
3	29.1	20	23.7	37	18.3
4	28.8	21	23.4	38	18.0
5	28.5	22	23.1	39	17.7
6	28.2	23	22.8	40	17.4
7	27.9	24	22.5	41	17.1
8	27.6	25	22.2	42	16.7
9	27.2	26	21.8	43	16.4
10	26.9	27	21.5	44	16.1
11	26.6	28	21.2	45	15.8
12	26.3	29	20.9	46	15.5
13	26.0	30	20.6	47	15.2
14	25.6	31	20.2	48	14.8
15	25.3	32	19.9	49	14.5
16	25.0	33	19.6	50	14.2

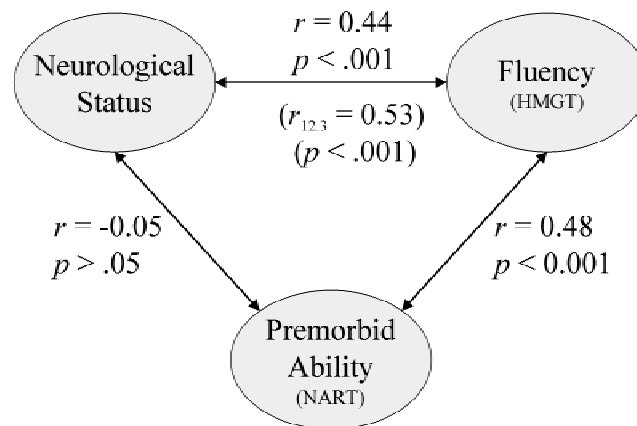


Fig. 1. Graphical illustration of the role of premorbid ability (measured by the NART) as a suppressor variable in the relationship between fluency and neurological status (the partial correlation between fluency and neurological status, controlling for premorbid ability, appears in brackets).

(the p value for this correlation is identical to the p value that would be obtained if an independent samples t test were used to compare the healthy and lesion samples). It can also be seen from Figure 1 that the NART is highly correlated with HMGT performance but does not correlate significantly with neurological status. Thus premorbid ability, as measured by the NART, fulfils the criteria for a suppressor variable. This was confirmed by computing the partial correlation between HMGT and group membership controlling for NART scores. This partial correlation (.53) is higher than the raw correlation (.44) between these variables. A method developed by Steiger (1980) was used to test whether, as hypothesized, the partial correlation was *significantly* higher than the raw correlation. This procedure tests the null hypothesis $\rho_{12} = \rho_{34}$, where, in the present case, 1 = HMGT, 2 = neurological status, 3 = the residuals obtained after predicting HMGT scores from the NART, and 4 = the residuals obtained after predicting neurological status from the NART (i.e., ρ_{12} is the raw correlation and ρ_{34} represents the partial correlation). This test revealed that the partial correlation was significantly higher than the raw correlation ($z = 2.85, p < .01$).

DISCUSSION

Measurement Characteristics of the HMGT

The Kolmogorov-Smirnov test revealed that HMGT raw scores did not depart significantly from a normal distribution. This property conveys on the HMGT the advantages identified by Warrington (2000) and Burgess and Shallice (1997). Not least among these is that HMGT scores can be analyzed using useful statistical methods that assume normality (including the methods employed in the present study).

The reliability of the HMGT was estimated using Cronbach's alpha. The alpha value obtained (.82) indicates that the HMGT has an acceptable level of reliability. Based on this alpha, confidence limits were obtained for individuals' scores on the HMGT. To illustrate the use and meaning of these confidence limits take the example of an individual who obtained a raw score of 17 on the HMGT. Using Warrington's (2000) table, this converts to a scaled score of 6. Consulting Table 2 it can be seen that the estimated true score is 7 and the accompanying lower and upper 95% confidence limits are 5 and 9 respectively. As noted, it is widely recommended that test scores should be accompanied by such limits as they serve the general purpose of reminding us that all test scores are fallible and they *quantify* the effects of this error. The confidence limits procedure produces limits on an individual's *true* score rather than *obtained* score; that is, there is a 95% probability that the individual's true score lies within these limits.

Estimation of Premorbid HMGT Performance

Correlational and factor analytic studies have demonstrated that the NART has high construct validity as a measure of

verbal intelligence (Crawford, 1992; O'Carroll, 1995). For example, a combined factor analysis of the NART and WAIS demonstrated that the NART loaded highly (.80) on the WAIS verbal factor (Crawford et al., 1989). Furthermore, Crawford et al. (2001) have recently reported a correlation of .73 between the NART scores of an elderly sample ($N = 179$) and the IQ scores this sample obtained in childhood (i.e., 66 years previously). In addition, NART performance has proved to be relatively unaffected by many neurological and psychiatric disorders; for example, see O'Carroll (1995) for a review.

Given this evidence, Warrington's (2000) report that the HMGT and the NART are highly correlated has two implications. Firstly, it indicates that, as is the case for other measures of verbal fluency, an individual's premorbid verbal IQ will partly determine performance on the HMGT. Secondly, it suggests that the NART can be used to control for the effects of premorbid verbal IQ when interpreting an individual's HMGT score. In the present study a regression equation was built to estimate premorbid HMGT scores from the NART. In clinical practice the estimated premorbid scores can then be compared with the scores obtained by patients on testing; a significant discrepancy in favor of the premorbid score would be taken as evidence for an acquired deficit.

Before turning to issues surrounding the validity of the equation, it is appropriate to briefly comment on the statistical method used to test for the significance of discrepancies. In the present study, critical values were obtained by multiplying the standard error of estimate by values of z . This method is widely used in clinical neuropsychology (e.g., Crawford et al., 1992; McSweeney et al., 1993; Paolo et al., 1996). However, numerous authorities on regression have pointed out that it is technically incorrect (Sokal & Rohlf, 1995; Zar, 1984). The correct method is to obtain the standard error of prediction for a new individual case (Cohen & Cohen, 1983); this standard error (rather than the standard error of the estimate) is then multiplied by a value of t (rather than a value of z) to obtain the critical values required. However, Crawford and Howell (1998b) compared the correct method with the approximate method used here in data simulated to represent a range of situations encountered in clinical neuropsychology. They reported that the technically incorrect method performs very well unless the sample size used to generate the equation is very small (not the case in the present study) and the scores on the predictor variable (NART in the present study) are *very* extreme. These conclusions have also subsequently been supported by examination of empirical data (Graves, 2000).

Validity and Utility of the Equation

The validity and utility of the NART regression equation was evaluated in three ways. As an initial validity check, the HMGT performance of anterior lesion cases was compared with the estimated premorbid scores provided by the regression equation. This yielded a highly significant dif-

ference ($t = 7.81, p < .001$). Although this result provides support for the *validity* of comparing obtained and estimated premorbid scores, it provides only limited support for its *utility*. It could be that simply using HMGT scores alone, rather than the discrepancy between premorbid and obtained scores, would be just as effective a means of detecting impairment. For example, an alternative regression-based means of estimating premorbid ability uses demographic variables as predictors (e.g., years of education, occupational status, age, etc). Eppinger et al. (1987) compared demographically estimated premorbid WAIS-R IQ scores with obtained IQs in a neurological sample and reported that premorbid IQs were significantly higher than the IQs obtained on testing. However, they also demonstrated that the discrepancies between premorbid and obtained IQs were no more effective than IQ scores alone at differentiating between the neurological cases and healthy controls.

Although there is now a substantial literature on the NART and its variants, this basic issue has received little empirical scrutiny. Crawford et al. (1990) used hierarchical discriminant function analysis to examine the ability of the NART in combination with WAIS IQ to correctly classify a sample consisting of healthy participants and patients with Alzheimer's disease (AD). The inclusion of the NART significantly improved the accuracy of classification over that achieved by WAIS IQs alone; 85% of cases were correctly classified by IQ scores and this rose to 96% for the combination of IQs and NART scores.

In the present study, the same methodology was employed to test whether the use of the NART to provide estimated premorbid scores for the HMGT would significantly improve its ability to discriminate between anterior lesion cases and healthy controls. The percentage of cases correctly classified rose from 75 to 81% when the NART was included in the discriminant function analysis and the change in classification accuracy was statistically significant. This result demonstrates that using the NART to provide an *individual* comparison standard for a patient's HMGT performance can supplement the use of conventional *normative* comparison standards.

The absolute percentage of cases correctly classified, either with or without the use of the NART, is relatively modest in comparison with Crawford et al.'s (1990) results. However, this is not unexpected for two reasons. Firstly, in the present study the presence of an anterior lesion is essentially used as a proxy for the presence of executive dysfunction. However, many patients with anterior lesions will not have suffered impairment of the executive system. Secondly, the executive system is complex and multifaceted. Therefore, it is entirely unrealistic to suppose that any single test will be able to detect all cases that have impairment of executive processes.

Premorbid IQ as a Suppressor Variable

Crawford et al. (1990) suggested that, in clinical populations, premorbid ability should be conceptualized as sup-

pressing the relationship between tests used to measure cognitive deficits and neurological status (regardless of whether neurological status is dichotomized, as in presence *versus* absence of a condition, or is continuous, as in an index of the severity of a condition). This suppression occurs because performance on virtually all neuropsychological tests will reflect, not only the effects of the presence or severity of neurological disease or trauma, but also pre-existing differences in ability. In contrast, premorbid ability will not, in general, be related to neurological status in conditions in which there is an adult onset. Therefore, there is a need to partial out (i.e., control for) the effects of premorbid ability so that our indices of test performance reflect impairment rather than an amalgam of variance attributable to impairment *and* premorbid ability.

The criterion for identifying whether a variable suppresses the relationship between two other variables is that the variable correlates significantly with one of the variables of interest but not with the other. It follows from this that controlling for the effects of the suppressor variable (i.e., partialling out its effect) will increase the magnitude of the correlation between the two other variables (Darlington, 1990; Howell, 1997). This is exactly the pattern obtained in the present study (see Figure 1).

It is also worth making explicit that, as the correlation between NART performance and neurological status is not significantly different from zero, the anterior lesion cases performed as well on the NART as the healthy sample; the p value for the point-biserial correlation between NART and group membership is identical to the p value that would be obtained if one conducted an independent samples t test on the NART scores of controls and anterior lesion cases. This result is important in its own right as, to our knowledge, there are no previous data on whether the NART and its variants can be used validly to estimate premorbid ability following focal frontal lesions. Investigation of this issue in other anterior lesion samples is warranted.

Crawford et al. (1990), in their study of the relationship between NART and WAIS IQ in healthy and AD samples, obtained exactly the same pattern of correlations as that observed in the present study. Thus these two studies, which employ different clinical samples (focal frontal lesions *vs.* AD) and different measures of current functioning (verbal fluency *vs.* WAIS IQ) provide converging evidence to support the rationale underlying the use of the NART in neuropsychological assessment.

Use of the Equation in Clinical Practice

The practicalities of using the present regression equation can be illustrated by the example of a 55-year-old male patient with a subdural hematoma in the left frontal lobe. His raw score on the HMGT was 19 (which converts to a scaled score of 7), and his error score on the NART was 12. Entering the NART error score into the regression equation produces an estimated premorbid score of 26.3 (see also Table 3). Therefore there is a discrepancy of 7.3 between

the estimated premorbid score and the obtained raw score. Consulting Table 4, it can be seen that this difference exceeds the critical value (6.4) for the .05 level of significance.

Critical values for more conservative significance levels (.15 and .10) are also provided in Table 4 because, inevitably, statistical power is low when working with individuals' scores. There is therefore the danger of committing Type II errors (i.e., wrongly rejecting the null hypothesis of "no deficit"). The choice of significance level is one for the clinician and will depend on the circumstances of the particular case; that is, the relative costs attached to false positives and false negatives.

Furthermore, as all significance levels are essentially arbitrary conventions, the provision of multiple critical values serves the additional, more general, purpose of allowing the clinician to estimate the abnormality of the discrepancy observed for their patient. Thus, if a patient's discrepancy falls between the critical values for the .10 and .05 levels, then the clinician knows that between 10 and 5% of the healthy population would be expected to obtain discrepancies that equal or exceed this discrepancy. If a more precise estimate of the abnormality of the discrepancy is required, then dividing the discrepancy by the equation's standard error of estimate (3.91) yields a z score which can then be referred to a table of the area under the normal curve (for the illustrative example discussed earlier, in which the discrepancy was 7.3, the z is 1.87 and the probability is therefore .03). Alternatively, the computer program that accompanies this paper can be used as it provides a precise probability for the discrepancy (see final section).

As noted, the regression equation developed in the present study is an example of the use of an individual comparison standard to detect and quantify the severity of cognitive deficits. It is notable that the performance of the patient in the illustrative example does not appear so extreme when compared against the HMGT normative data. The scaled score on the HMGT was 7 which corresponds to a z score of exactly -1.0 (scaled scores have an SD of 3). Referring this z value to a table of the area under the normal curve¹ yields a one-tailed probability of .16. A scaled score of 5 (i.e., a score that was more than 1.64 SD units below the mean) would be required for significance at the .05 level. This reinforces Lezak's (1995) emphasis on the usefulness of *individual* comparison standards when attempting to identify neuropsychological deficits.

The present study used the NART as a means of estimating premorbid fluency ability. However, variants on this test have been developed specifically for use in the United

States and Canada; that is, the North American Reading Test (NAART; Blair & Spreen, 1989) and the American National Adult Reading Test (AMNART; Grober & Sliwinski, 1991). Given that the present results were positive, it would be worth developing and evaluating equations based on these variants for use with both conventional verbal fluency tests and the HMGT.

Computer Program for the HMGT

The labor involved in using the methods provided in the present paper is modest. However, we considered it would be more convenient for clinicians if the procedures were implemented in a computer program for PCs (this should also minimize the risk of clerical error). The program takes a patient's raw score on the HMGT and (optionally) their NART error score. The output consists of the HMGT scaled score, the estimated true score, and the 95% confidence limits on true scores. It also converts NART error scores to estimated premorbid HMGT scores and reports the one-tailed probability for the discrepancy between the estimated premorbid score and the score obtained on testing. The program can be downloaded from the first author's website (Crawford, 2002).

REFERENCES

- Blair, J.R. & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist*, 3, 129–136.
- Borkowski, J.G., Benton, A.L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, 5, 135–140.
- Burgess, P.W. & Shallice, T. (1997). *The Hayling and Brixton Tests. Test manual*. Bury St Edmunds, UK: Thames Valley Test Company.
- Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Crawford, J.R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J.R. Crawford, D.M. Parker, & W.W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 21–49). London: Erlbaum.
- Crawford, J.R. (1996). Assessment. In J.G. Beaumont, P.M. Kenaley, & M.J. Rogers (Eds.), *The Blackwell dictionary of neuropsychology* (pp. 108–116). London: Blackwell.
- Crawford, J.R. (2002). A computer program for analysis of scores on the Homophone Meaning Generation Test. WEB: <http://www.psyc.abdn.ac.uk/homedir/jcrawford/HMGT.htm>
- Crawford, J.R., Deary, I.J., Starr, J.M., & Whalley, L.J. (2001). The NART as an index of prior intellectual functioning: A retrospective validity study covering a 66 year interval. *Psychological Medicine*, 31, 451–458.
- Crawford, J.R., Hart, S., & Nelson, H.E. (1990). Improved detection of cognitive impairment with the NART: An investigation employing hierarchical discriminant function analysis. *British Journal of Clinical Psychology*, 29, 239–241.
- Crawford, J.R. & Howell, D.C. (1998a). Comparing an individual's test score against norms derived from small samples. *Clinical Neuropsychologist*, 12, 482–486.

¹The use of z for this purpose is widespread (Howell, 1997; Ley, 1972) but it should be noted that the method treats the normative sample mean and SD as though they were population means and SD s. Crawford and Howell (1998a) describe a more valid method that treats the sample statistics as sample statistics (rather than population parameters) and uses the t distribution to evaluate the rarity/statistical significance of the score. However, although the use of z is inappropriate when N for the normative sample is small, it yields results that are, for all practical purposes, indistinguishable from the t distribution method when the normative sample is as large as the present one.

- Crawford, J.R. & Howell, D.C. (1998b). Regression equations in clinical neuropsychology: An evaluation of competing methods for comparing predicted and obtained scores. *Journal of Clinical and Experimental Neuropsychology*, *20*, 755–762.
- Crawford, J.R., Moore, J.W., & Cameron, I.M. (1992). Verbal fluency: A NART-based equation for the estimation of premorbid performance. *British Journal of Clinical Psychology*, *31*, 327–329.
- Crawford, J.R., Obonsawin, M.C., & Bremner, M. (1993). Frontal lobe impairment in schizophrenia: Relationship to intellectual functioning. *Psychological Medicine*, *23*, 787–790.
- Crawford, J.R., Stewart, L.E., Cochrane, R., Parker, D.M., & Besson, J.A.O. (1989). Construct validity of the National Adult Reading Test: A factor analytic study. *Personality and Individual Differences*, *10*, 585–587.
- Crawford, J.R., Venneri, A., & O'Carroll, R.E. (1998). Neuropsychological assessment of the elderly. In A.S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology*, Vol. 7: *Clinical geropsychology* (pp. 133–169). Oxford, UK: Pergamon.
- Darlington, R.B. (1990). *Regression and linear models*. New York: McGraw-Hill.
- Eppinger, M.G., Craig, P.L., Adams, R.L., & Parsons, O.A. (1987). The WAIS-R index for estimating premorbid intelligence: Cross-validation and clinical utility. *Journal of Consulting and Clinical Psychology*, *55*, 86–90.
- Feldt, L.S. (1965). The approximate sampling distribution of Kuder-Richardson Reliability Coefficient Twenty. *Psychometrika*, *30*, 357–370.
- Franzen, M.D. (1989). *Reliability and validity in neuropsychological assessment*. New York: Plenum Press.
- Franzen, M.D., Burgess, E.J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology*, *12*, 711–738.
- Glutting, J.J., McDermott, P.A., & Stanley, J.C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement*, *1987*, 607–614.
- Grant, D.A. & Berg, E.A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*, 404–411.
- Graves, R.E. (2000). Accuracy of regression equation prediction across the range of estimated premorbid IQ. *Journal of Clinical and Experimental Neuropsychology*, *22*, 316–324.
- Grober, E. & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, *13*, 933–949.
- Heaton, R.K. (1981). *Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Howell, D.C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.
- Ley, P. (1972). *Quantitative aspects of psychological assessment*. London: Duckworth.
- Lezak, M.D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.
- McCarthy, R.A. & Warrington, E.K. (1990). *Cognitive neuropsychology: A clinical introduction*. San Diego, CA: Academic Press.
- McKenna, P. & Warrington, E.K. (1983). *Graded Naming Test manual*. Windsor, UK: NFER-Nelson.
- McSweeney, A.J., Naugle, R.I., Chelune, G.J., & Lüders, H. (1993). "T Scores for Change": An illustration of a regression approach to depicting change in clinical neuropsychology. *Clinical Neuropsychologist*, *7*, 300–312.
- Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. *British Journal of Clinical Psychology*, *23*, 53–57.
- Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*, 313–324.
- Nelson, H.E. & Willison, J. (1991). *National Adult Reading Test manual* (2nd ed.). Windsor, UK: NFER-Nelson.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Carroll, R. (1995). The assessment of premorbid ability: A critical review. *Neurocase*, *1*, 83–89.
- Paolo, A.M., Ryan, J.J., Tröster, A.I., & Hilmer, C.D. (1996). Demographically based regression equations to estimate WAIS-R subtest scaled scores. *Clinical Neuropsychologist*, *10*, 130–140.
- Silverstein, A.B. (1989). Confidence intervals for test scores and significance tests for test score differences: A comparison of methods. *Journal of Clinical Psychology*, *45*, 828–832.
- Sokal, R.R. & Rohlf, J.F. (1995). *Biometry* (3rd ed.). San Francisco: W.H. Freeman.
- Stanley, J.C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 356–442). Washington DC.: American Council on Education.
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.
- Tabachnick, B.G. & Fidell, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
- Warrington, E.K. (2000). Homophone meaning generation: A new test of verbal switching for the detection of frontal lobe dysfunction. *Journal of the International Neuropsychological Society*, *6*, 643–648.
- Zar, J.H. (1984). *Biostatistical analysis* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.