Journal of the International Neuropsychological Society (2000), **6**, 643–648.

Homophone meaning generation: A new test of verbal switching for the detection of frontal lobe dysfunction

ELIZABETH K. WARRINGTON
Dementia Research Group, Institute of Neurology, London, UK

(Received March 25, 1999; Revised July 15, 1999; Accepted July 20, 1999)

Abstract

Executive functions in their broadest sense may be impaired in patients with frontal lobe lesions. The Wisconsin Card Sorting Test, perhaps the most robust test sensitive to frontal lobe dysfunction, requires the flexibility to switch a response pattern to meet the change in task demands. However this task has only two category switches and normal respondents tend to score at ceiling. Verbal fluency tasks also incorporate a switching component and it was hypothesized that a word generation task that maximized a switching requirement might provide a more satisfactory verbal measure of frontal lobe dysfunction. A new test of homophone meaning generation that requires multiple switches between verbal concepts (e.g., *tick* = insect, correct, etc.) was devised. Normative data was obtained from a sample of 170 control participants. Seventy-one patients with unilateral anterior or posterior cerebral lesions were also tested. A normal distribution of scores was obtained in the standardization sample. The anterior lesion groups were more impaired than the posterior groups. There were no significant differences due to laterality. This homophone meaning generation task is a measure of frontal lobe dysfunction that has the advantage of psychometric properties that permit measurement of the degree of impairment. (*JINS*, 2000, *6*, 643–648.)

Keywords: Verbal fluency, Switching, Frontal lobes, Homophone meaning generation

INTRODUCTION

Frontal lobe dysfunction has been characterized in many ways. A failure of the abstract attitude, control mechanisms, modulatory functions, problem solving, planning and strategy formulation and supervisory attentional systems have all been proposed to encapsulate frontal lobe deficits. Executive functions in their broadest sense would appear to be impaired. But even the most unitary account must allow for there being subcomponents of the system. It is therefore hardly surprising that reliable tests to detect frontal lobe dysfunction have proved very elusive.

Of all the tests incorporated in batteries to assess frontal lobe dysfunction, perhaps one of the most robust is the test that has come to be called the Wisconsin Card Sorting Test. In some studies a selective right frontal deficit has been observed (Robinson et al., 1980), in others a left frontal deficit (Drewe, 1974; Milner, 1964; Taylor, 1979), and in still others both a right and a left frontal deficit have been found (e.g., Nelson, 1976). One crucial aspect of the test would appear to be the necessity to have the flexibility to switch a response pattern to meet the change in the task demands. The Wisconsin Card Sorting Test has the possible advantage of being nonverbal but the disadvantage of having only three categories with five switches. Normal participants tend to obtain a ceiling score; thus there is no adequate scale against which to assess an impaired score other than to classify it as a failure. Its psychometric properties, in common with many other tasks (other than those tests that incorporate a time measure) held to be sensitive to frontal lobe pathology, are limited in that they only yield a pass or fail measure.

Verbal fluency tests, which are commonly used as a measure of frontal lobe dysfunction, are not subject to this limitation. There are several problems with such tasks. It is inappropriate for patients with word retrieval or speech production difficulties. The strict timing conditions can be a confounding factor. In a clinical setting fluency tasks can be adversely affected by anxiety. Recently, a more detailed analysis of fluency tasks have shown them to comprise two components: *clustering* and *switching*. Moreover, it has been argued that it is only the latter component, that of switching, that is compromised by frontal lobe dysfunction (e.g., Chertkow & Bub, 1990; Gruenewald & Lockhead, 1980; Troyer et al., 1997).

Reprint requests to: Professor Elizabeth Warrington, Dementia Research Group, National Hospital, Queen Square, London, WC1N 3BG, UK.

https://doi.org/10.1017/S1355617700666018 Published online by Cambridge University Press

A task that maximized the switching component in a word generation task might provide a more satisfactory verbal measure of frontal lobe dysfunction than is at present available. There are many homophones in the English language. Many of these homophones have multiple high frequency meanings (e.g., *form*, *slip*). Clinical observation suggested that the ability to generate multiple meanings could provide a means of assessing the ability to switch between alternative verbal concepts. Furthermore, by using a pool of homophonic words, there was clearly the potential to achieve a test with scores that were normally distributed.

Our aim in this investigation was to construct, standardize, and validate a test of verbal switching. First we describe a new task of homophone meaning generation that requires multiple switches between verbal concepts. Second we describe our standardization study, in which a large representative cross section of an urban population have been tested. Third we report a validation study in which we demonstrate that patients with unilateral frontal lobe lesions are more impaired on the verbal switching task than patients with more posterior cerebral damage.

CONSTRUCTION AND STANDARDIZATION OF THE WORD-MEANING GENERATION TEST

Research Participants

Our aim was to test a representative cross section of the population attending public sector London hospitals. The standardization sample consisted of 170 volunteers (68 male and 102 female) aged from 19 to 74 years ($M = 45.1$, $SD = 14.9$). The participants came from several sources, including employees of London companies (e.g., The Prudential), attendants at day centers, and friends and relatives accompanying out-patients at the National Hospital. The spread of economic status was satisfactorily wide ranging from unskilled manual workers to those with a professional qualification. The largest number (80 participants) were in the C1 category which includes skilled, nonprofessional individuals. All were English speaking and had been educated in the British educational system (minimum 10 years).

Test Construction

A pool of homophonic words were selected, eight of them with a single spelling (e.g., *slip*, *form*) and eight with more than one spelling (e.g., *sent*, *scent*). All the words had at least three common meanings. The words were presented orally and participants were instructed to indicate either by definition or by gesture as many different meanings as possible for each word presented. No time constraints were imposed and the participant usually indicated their readiness to proceed to the next stimulus word. One point was awarded for each distinct meaning generated for each word. A consensus between four scorers regarding what constituted a distinct meaning was achieved. In order to shorten the test for clinical purposes eight words (four same spelling: *form*, *slip*, *tick*, and *tip*, and four different spelling: *bear*, *cent*, *right*, and *bored*) from the original pool of 16 were selected on the basis that there was the greatest spread of scores in the standardization sample. A score for the four same-spelling words, the four different-spelling words, and a total score for all eight words was recorded. The following analysis is based on this shortened version of the test.

Procedure

The participants were tested individually in one session. In addition to the new homophone meaning generation test, two baseline tests were administered:

1. The National Adult Reading Test (NART). This test was included as an estimate of intellectual functioning in the standardization sample and to provide further evidence that a representative cross-section of the population had been sampled (Nelson & Willison, 1992).
2. The Graded Naming Test (GNT). This naming test was included to provide a second measure of verbal vocabulary in the standardization sample (McKenna & Warrington, 1983).

RESULTS

Baseline Tests

The mean and standard deviation of the raw scores of the standardization sample on the NART and the GNT are given in Table 1. The mean reading IQ equivalent of this sample estimated from the mean NART error score is 100 and the mean GNT score in this population is very close to that obtained in a recent restandardization for which the mean GNT score was 22.5 (Warrington, 1997).

Table 1. Standardization sample: Mean age and test scores

Parameter	Age	NART	GNT	Spelling same	Spelling different	Total score
M	45.1	22.9	20.8	11.8	11.9	23.7
SD	14.9	9.3	4.1	3.1	2.3	4.9
Range	19–74	3–50	7–28	4–19	6–18	10–35

Homophone Meaning Generation Test

The mean and standard deviation of the scores on each section of the word meaning generation test together with the mean total score and standard deviation are given in Table 1. The correlation between age and performance on the word generation test (total score) covarying for performance on the NART was not significant ($r = 0.12$, $p > .2$). Consequently the 170 participants of this study were not subdivided into age bands and the further analysis treats them as a single group.

The scores on the same-spelling and the different-spelling homophones were not significantly different (related $t = 0.64$, $p > .2$). The correlation between the NART and the homophone meaning generation test was highly significant ($r = .6$, $p < .0001$). The correlation between the homophone meaning generation test and the performance on the GNT covarying for performance on the NART was $r = .22$ ($p < .01$). Percentile scores for the three measures (*same*, *different*, and *total*) of the test are given in Appendix A and normalized scores ($M = 10$, $SD = 3$) are given in Appendix B.

COMMENT

This standardization study is satisfactory insofar as a representative cross-section of the population has been tested. Their mean estimated IQ from the NART is very close to the mean of the original population standardization. No significant age effects have been obtained, which suggests that this test may have features in common with other measures of "crystallized" intelligence that are relatively unaffected by age. The test was administered in two sections, generating meanings for the same-spelling and different-spelling homophones. The fact that there was no difference between an individual's performance generating meanings for these two types of homophone indicates that the reliability of the task is very satisfactory and could also provide parallel forms of the test.

VALIDATION STUDY

The comparison of groups of patients with unilateral cerebral lesions on a particular task allows one to assess its pragmatic strength in terms of localization and lateralization specificity. More specifically the aim in this validation study was to establish whether patients with unilateral frontal lobe lesions would be more impaired on the homophone task than patients with unilateral lesions that spared the frontal cortices and in addition to establish whether performance on this highly verbal task would be related to the laterality of the lesion.

Research Participants

A consecutive series of 71 patients with unilateral cerebral lesions of recent onset were tested. All patients physically fit enough to be tested in the psychology department of the National Hospital and able to cooperate with the task demands were included (dysphasia was not a basis for exclusion). Patients who had undergone a temporal lobectomy in treatment for long-standing intractable epilepsy were excluded as they were deemed to have had long-standing lesions. Otherwise the same exclusion criteria as for the standardization study were applied. Fifty of these patients had space occupying tumors and the majority of the remaining 21 cases had a well-localized vascular lesion. On the basis of CT scan or MRI scan evidence, each patient was allocated to one of four localization groups as follows:

1. *Left anterior* ($N = 17$): patients with left frontal, left frontoparietal, or left frontotemporal lesions.
2. *Left posterior* ($N = 20$): patients with left parietal, left temporal, left temporoparietal, occipitoparietal, or left occipitotemporal lesions.
3. *Right anterior* ($N = 19$): patients with right frontal, right frontoparietal, or right frontotemporal lesions.
4. *Right posterior* ($N = 15$): right parietal, right temporal, right temporoparietal, right occipitoparietal, or right occipitotemporal lesions.

Patients with large lesions either extending across the midline or involving three or more cerebral regions were excluded.

Procedure

Patients were tested individually in one session. The testing procedure differed in two respects from that adopted in the standardization study:

1. *The Modified Wisconsin Card Sorting Test* (Nelson, 1976) was administered (in addition to the two baseline tests GNT and NART included in the standardization study) to provide a conventional measure of frontal lobe dysfunction. The number of categories sorted (maximum six) and the total number of errors were recorded.
2. *Homophone Meaning Generation Test*: The shortened version of the test was administered. The four same-spelling homophones were attempted before the four different-spelling homophones.

RESULTS

The mean age and scores on the baseline tests for each of the localization subgroups and the two laterality groups are given in Table 2. A Laterality × Location ANOVA was computed for each measure. The only finding of note was a significant anterior location effect on the categories measure of the Wisconsin Card Sorting Test (see Table 3). Subsequent analysis indicated that the right and left anterior groups were not significantly different ($t = 0.4$, $p > .5$).

https://doi.org/10.1017/S1355617700666018 Published online by Cambridge University Press

Table 2. Mean ages and mean scores on the baseline tests for each lesion group

	Age		NART		GNT		WCST categories		WCST errors	
Lesion group	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Left anterior (*N* = 17)	48.35	(12.9)	23.4	(10.5)	18.2	(5.3)	4.3	(1.9)	11.3	(8.6)
Left posterior (*N* = 20)	45.9	(13.8)	23.3	(10.9)	19.1	(5.7)	5.1	(1.7)	9.8	(11.3)
Right anterior (*N* = 19)	44.8	(15.6)	20.2	(8.2)	18.1	(5.4)	4.0	(2.4)	13.0	(12.1)
Right posterior (*N* = 15)	47.3	(14.0)	20.2	(8.7)	19.8	(5.0)	5.5	(1.7)	10.4	(10.4)
Left hemisphere (*N* = 37)	45.9	(13.8)	23.3	(10.9)	18.7	(5.4)	4.7	(1.8)	10.4	(10.1)
Right hemisphere (*N* = 34)	46.2	(14.6)	20.2	(8.7)	19.8	(5.0)	4.7	(2.2)	10.4	(10.7)

Table 3. Lesion group comparisons of baseline test scores (ANOVA)

	NART		GNT		WCST categories		WCST errors	
Groups	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Laterality	1.6	>.2	.8	>.2	0.2.	>.5	0.01	>.5
Location	0.0	>.5	2.7	>.1	6.4	<.02	3.6	<.1
Laterality × Location	0.0	>.5	1.0	>.1	0.5	>.5	0.4	>.5

Table 4. Mean scores and standard deviations on the Homophone Meaning Generation Test for each lesion group

	Spelling same		Spelling different		Total score		Perseverative responses
Lesion group	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	(mean no.)
Left anterior (*N* = 17)	8.2	(3.3)	8.5	(3.0)	16.7	(3.1)	1.2
Left posterior (*N* = 20)	10.6	(3.1)	11.2	(3.0)	21.8	(5.7)	0.75
Right anterior (*N* = 19)	8.4	(2.8)	8.7	(3.2)	17.1	(5.8)	4.6
Right posterior (*N* = 15)	12.3	(3.1)	11.8	(2.0)	24.1	(4.5)	1.4
Left hemisphere (*N* = 37)	9.5	(2.8)	10.0	(3.0)	19.4	(5.7)	1.0
Right hemisphere (*N* = 34)	10.2	(3.2)	10.2	(3.2)	20.4	(6.3)	3.1

Table 5. Lesion group comparisons of homophone meaning generation scores (ANOVA)

	Spelling same		Spelling different		Total score	
Groups	*F*	*p*	*F*	*p*	*F*	*p*
Laterality	0.4	>.5	1.9	>.1	1.2	>.2
Location	19.2	<.000	20.2	<.000	23.25	<.000
Laterality × Location	0.1	>.5	1.0	>.2	0.6	>.2

https://doi.org/10.1017/S1355617700666018 Published online by Cambridge University Press

The mean scores and standard deviations of the three homophone meaning generation measures (*same* spelling, *different* spelling, and total score) are given in Table 4. Laterality by location ANOVAs were computed (see Table 5). The effect of location was significant on all three measures, indicating an anterior deficit. There was no effect of laterality and the interaction terms were not significant. There were no significant differences between the same-spelling homophones and the different-spelling homophones in either location groups (anterior: $t = -.86$; posterior: $t = .29$). Subsequent analyses using the total generation score demonstrated that the left anterior group were significantly more impaired than the left posterior group ($t = 2.97$, $p < .005$) and the right anterior group were more impaired than the right posterior group ($t = 3.8$; $p < .001$). However there were no significant differences between the right and left anterior groups ($t = 0.2$, $p < .5$).

Qualitatively, it was noted that some individuals gave examples of using the same core meaning of the homophonic word (e.g., *slip*, falling over, accident, skidding, sliding). The number of perseverative responses was recorded for each participant (see Table 4). A Mann–Whitney *U* test was computed to compare the laterality and location groups. There was a trend for the right hemisphere group to make more perseverative responses than the left hemisphere group ($z = 1.84$, $p < .1$) and for the anterior lesion group to make more of this type of error than the posterior lesion group ($z = 1.72$, $p < .1$). Thus, although in terms of total number of homophone meanings generated there was no difference between the right and left anterior groups, there was a definite trend for the right anterior group to make more perseverative responses.

The number of individuals failing the Wisconsin Card Sorting Test test (by the criterion of less than five categories) and failing the homophone test (scoring below the 5% cut-off) was recorded. The coefficient of concordance between the two tasks was significant ($\chi^2 = 4.25$, $p < .05$).

DISCUSSION

The first aim of this investigation, to standardize, in a representative cross-section of the population, a graded difficulty test of verbal switching, has been met. The new test of homophone meaning generation was devised to focus on the requirement to switch between verbal concepts. In the English language there are many homophonic words that have totally distinct meanings. For example consider the word *form*; considerable conceptual switches are required to generate *a shape*, *a piece of paper*, *a school class*, and *a bench*. In order to minimize possible confounding effects of a poor vocabulary or comprehension loss the homophonic words selected had multiple high-frequency meanings.

For our normal sample there was no greater difficulty in generating multiple meanings for homophones with the same spelling (e.g., *form*, *slip*) as with different spellings (e.g., *right*, *write*). This result provides the possibility to use the two sections of the test independently as parallel forms. Not only were the combined total scores on the test normally distributed about the mean, this was also the case for each section of the test; both floor and ceiling effects have been avoided. Thus it has the psychometric properties of other graded difficulty tests of intellectual and cognitive skills. It is therefore valid to normalize the scores which then permits direct comparison of performance on this test with measures of other cognitive skills. The range of scores, especially the combined total score on the test, is sufficiently great that it also has the power to monitor improvement and deterioration in an individual's ability to do this task.

The second aim of this investigation was to establish whether there would be a selective impairment in patients with unilateral anterior lesions. The results of the validation study were clear-cut. The anterior lesion groups were significantly more impaired than the posterior lesion groups, whereas there were no significant laterality effects. A more detailed *post-hoc* analysis established that both the right and left frontal lesion groups were impaired compared with their comparison posterior lesion groups. Despite this being a transparently verbal test the right anterior lesion group were as impaired as the left anterior group. Thus we have obtained a bilateral anterior lesion deficit on this homophone meaning generation task. A similar pattern of results was obtained for the category measure on the Wisconsin Card Sorting Test: there was a significant effect of localization but not for lateralization. An absence of a laterality effect has also been reported with other executive tasks sensitive to frontal lobe dysfunction (e.g., cognitive estimates; Shallice & Evans, 1978).

The question arises as to whether the basis of the deficit differs in the two laterality groups and that this task too, as has already been shown for fluency tests, comprises multiple components. The observation that some individuals were able to give numerous examples of a single meaning (e.g., *tip* of a pen, *tip* of a mountain, *tip* of an iceberg, etc.) but could not switch to a different concept might point to there being a distinction between perseverative responding and a failure to switch. That this type of error was somewhat more frequent in the right anterior group gives some credence to this distinction. It would appear that both these operations may have an anterior localization, but lateralization of the deficit is only observed for perseverative responses.

Fluency tests of word generation have been shown to have two major components: clustering, that of generating words within a semantic or phonemic subcategory; and switching, the ability to switch between clusters. The homophone meaning generation task ostensibly focuses on one component of a fluency task, the switching component. But here too it would appear that there may be multiple components. The ability to inhibit inappropriate items from the same cluster, resulting in perseverative responses, appears to dissociate from the ability to switch to a new cluster. By the same token seemingly unidimensional executive-type tasks might yield to a more fine-grained analysis and thus resolve some of the conundrums of the anatomical correlates of the dysexecutive syndromes.

https://doi.org/10.1017/S1355617700666018 Published online by Cambridge University Press

REFERENCES

Chertkow, H. & Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type. *Brain*, *113*, 397–417.

Drewe, E.A. (1974). The effect of type and area of brain lesion on Wisconsin Card Sorting Test performance. *Cortex*, *10*, 159–170.

Gruenewald, P.J. & Lockhead, G.R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, *25*, 225–240.

McKenna, P. & Warrington, E.K. (1983). *The Graded Naming Test*. Windsor, UK: Nelson.

Milner, B. (1964). Some effects of frontal lobectomy in man. In J.M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 122–145). New York: McGraw-Hill.

Nelson, H.E. (1976). A modified card sorting task sensitive to frontal lobe defects. *Cortex*, *12*, 313–324.

Nelson, H.E. & Willison, J.R. (1992). *Re-standardization of the NART against the WAIS–R*. Windsor, UK: Nelson.

Robinson, A.L., Heaton, R.K., Lehmann, R.A.W., & Stilson, D.W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lesions. *Journal of Consulting and Clinical Psychology*, *48*, 605–614.

Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*, 294–303.

Taylor, L. (1979). Psychological assessment of neurosurgical patients. In T. Rasmussen & R. Marino (Eds.), *Functional neurosurgery* (pp. 165–180). New York: Raven.

Troyer, A.K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*, 138–146.

Warrington, E.K. (1997). The Graded Naming Test: A restandardization. *Neuropsychological Rehabilitation*, *7*, 143–146.

APPENDIX A

Percentile scores

Percentile score	Spelling same	Spelling different	Total score
5	6	8	14
10	7	9	17
25	10	11	21
50	12	12	24
75	14	13	27
90	16	15	30
95	16.5	16	31

APPENDIX B

Conversion table of raw scores to scaled scores

Scaled score	Spelling same	Spelling different	Total score
1	2	5	9
2	3	6	10–11
3	4–5	7	12–13
4	6	—	14
5	7	8	15–16
6	8	9	17–18
7	9	10	19
8	10	—	20–21
9	11	11	22
10	12	12	23–24
11	13	13	25–26
12	14	—	27
13	15	14	28–29
14	16	15	30–31
15	17	16	32
16	18	—	33–34
17	19	17	35
18	20	18	36–37
19	21	19	38–39
20	22	—	40