The Modified Fuld Verbal Fluency Test: A Validation Study in Hong Kong

Helen F. K. Chiu,1 Calais K. Y. Chan,1 Linda C. W. Lam,1 Kee-on Ng,1 Siu-wah Li,2 Mary Wong,2 and Wah-fat Chan3

1Department of Psychiatry, Chinese University of Hong Kong.  
2Castle Peak Hospital, Hong Kong.  
3Pamela Youde Nethersole Eastern Hospital, Hong Kong.

Recent studies in the West have found that verbal fluency tests are useful in the detection of dementia. In this study, the Fuld Verbal Fluency Test (FVFT) was modified and tested for its usefulness as a screening instrument for differentiating demented from normal elderly Chinese in Hong Kong. Fifty-three normal and 56 demented subjects were administered the FVFT, the Cantonese version of the Mini-Mental State Examination (CMSE), and the Chinese version of the Hamilton Depression Rating Scale. In addition to bearing significant correlation with the CMSE, the verbal fluency scores for all semantic categories (animals, fruits, vegetables) were found to produce significant differentiation between the normal and demented groups. However, test-retest reliability for the individual verbal fluency scores was shown to be unsatisfactory, and a composite measure was recommended for use in future screening on the grounds of its enhanced reliability.

SEVERAL controlled studies have attested to the use of verbal fluency (VF) tests in the detection of dementia (Hart, Smith, & Swash, 1988; Monsch et al., 1992; Ober et al., 1986). Two common forms of VF are the letter and category fluency. While the former typically requires subjects to generate words that begin with a given letter, the latter requires subjects to produce exemplars of a given category (e.g., animals) within a designated time limit.

Recently, differential characteristics in test performance on VF tests have been found in specific types of dementia (Hanley, Dewick, Davies, Pfafer, & Turnbull, 1990; Monsch et al., 1994). For instance, while category fluency has been found to be disproportionately more impaired than letter fluency in patients with Alzheimer’s disease, performances on category and letter fluency have been observed to be equally impaired among patients of Huntington’s and Parkinson’s diseases (Butters, Wolfe, Granholm, & Martone, 1986; Hodges, Salmon, & Butters, 1990). It is generally interpreted that such a differential pattern of performance indicates specific impairments in access and/or store of semantic knowledge in Alzheimer’s patients as compared to other forms of dementia. Moreover, the use of phonemic and lexical cues by dementia patients has been suggested to preserve relatively better performance on the letter fluency tasks.

As yet, there are very few tests validated for use in the assessment of dementia among Chinese in Hong Kong. The most commonly used test has been the Cantonese version of the Mini-Mental State Examination (Chiu, Lee, Chung, & Kwong, 1994). In our attempt to widen the scope of screening tools for dementia, the use of letter fluency tests is discouraged by the distinctive linguistic properties of Chinese, which deviate markedly from the Western alphabet system. On the other hand, the application of category fluency tests does not seem to lend itself readily to such intrinsic problems. Hence, we sought to examine performance of both normal and demented elderly subjects on the Fuld Verbal Fluency Test (FVFT), a typical category fluency test adopted from the Fuld Object Memory Test (Fuld, 1977). In particular, the present study aimed to evaluate the usefulness of FVFT as a screening instrument for differentiating dementia from normal cases in our locality. This is part of a more extensive project on the development of a comprehensive neuropsychological battery for Chinese patients with dementia in Hong Kong.

METHOD

Subjects
Two groups of subjects were included in the study. The first consisted of normal elderly subjects recruited from a local social center for the elderly and an old age home (n = 53). The second group consisted of demented subjects identified by visiting psychogeriatricians in the old age home, as well as demented patients referred to a university psychogeriatric unit over a 6-month period (n = 56). Demented subjects were diagnosed according to the DSM-IV criteria (American Psychiatric Association, 1994). Patients with profound sensory impairment, significant physical disability, and identified neurological and psychiatric disorders other than dementia were excluded from the study. Among the 56 demented subjects, 40 had Alzheimer’s disease, 13 vascular dementia, and 3 had dementia due to other causes. All subjects consented to participate in the study.

The mean age of the 109 subjects was 75.8 ± 7.9 (range = 59–97), and the mean age of demented subjects (77.4 ± 8.5) was slightly higher than that of normal controls (74.1
was slightly higher among the dementia subjects, 4.1 ± 2.7, as compared to the normal controls, 3.1 ± 3.7; however, this failed to reach statistical significance \[F(1,106) = 2.59, p > .05]\.

**Verbal Fluency Scores**

The mean scores for the categories of "animals," "fruits," "vegetables," "things that make you happy," and "things that make you sad" for the normal subjects were 12.4 ± 4.4, 8.3 ± 2.4, 9.0 ± 3.0, 3.2 ± 1.9, and 1.8 ± 1.4, respectively. For the demented subjects, the corresponding figures were 5.9 ± 3.3, 3.3 ± 2.3, 4.9 ± 3.1, 1.7 ± 1.4, and 1.0 ± 0.9, respectively. For both groups, the largest number of items reported for the category "animals" was clearly a result of the longer test duration (60 s). On the other hand, the number of items reported for the "fruits" and "vegetables" categories was significantly more than that of the mood-related categories despite the same test duration (30 s). Moreover, the number of items reported for "things that make you happy" was significantly more than the number of items reported for "things that make you sad" \[F(1,106) = 53.72, p < .05]\, particularly among the control subjects as revealed by a significant group interaction with the mood-related categories \[F(1,105) = 4.45, p < .05]\.

Nevertheless, all of the VF scores produced reliable differences between the two groups at \(p < .01\). The \(F\) coefficients for the categories "animals," "fruits," and "vegetables" were \(F(1,106) = 79.46, F(1,106) = 126.00,\) and \(F(1,106) = 49.42,\) respectively, whereas the \(F\) coefficients for the positive and negative mood-related categories were \(F(1,106) = 19.56\) and \(F(1,106) = 13.19,\) respectively. Based on these results, a combined VF score across the categories "animals," "fruits," and "vegetables" (CVF) was included in the analysis to assess its additional power in differentiating the two groups. The CVF was 14.0 ± 7.9 for the demented subjects and 29.7 ± 8.2 for the normal subjects, and it yielded good differentiation between groups \[F(1,106) = 101.91, p < .05\].

**Validity and Reliability**

In regard to convergent validity, all of the VF scores bore significant correlation with the CMMSE score, \(r = .64\) for "animals" \((p < .01)\), \(r = .47\) for "fruits" \((p < .01)\), \(r = .44\) for "vegetables" \((p < .01)\), \(r = .59\) for "things that make you happy" \((p < .01)\), and \(r = .33\) for "things that make you sad" \((p < .01)\). In regard to test-retest reliability, all of the VF scores produced only fair outcomes, and the test-retest correlations obtained were, \(r = .67\) for "animals" \((p < .05)\), \(r = .45\) for "fruits" \((p > .05)\), \(r = .58\) for "vegetables" \((p < .05)\), \(r = .68\) for "things that make you happy" \((p < .05)\), and \(r = .31\) for "things that make you sad" \((p > .05)\). By contrast, the CVF was found to yield the best correlation with the CMMSE, \(r = .76\) \((p < .01)\), and the test-retest correlation was also more satisfactory, \(r = .74\) \((p < .01)\).

**Sensitivity and Specificity**

Table 1 shows the optimal VF cut-off scores for the categories of "animals," "fruits," "vegetables," and the CVF, as well as their corresponding rates of sensitivity and specificity. For the individual categories, the respective cut-off scores were 9 for "animals," 6 for "fruits," and 7 for "vegetables," whereas a cut-off score of 19 apparently produced the best outcome for the CVF.
Effects of Age, Sex, and Education

Age was found to be inversely correlated with VF scores of “animals” \( (r = -.29, p < .05) \) and “fruits” \( (r = -.34, p < .05) \), but not with “vegetables” \( (r = -.19, p > .05) \). While gender produced no significant effect, \( F(1,106) < 4.7, p > .05 \), the VF scores were also found to be uncorrelated with educational levels \( (r < .15, p > .05) \). In addition, there were no significant interaction effects pertaining to age, sex, and education, \( F < 3, p > .05 \).

DISCUSSION

This study was aimed to evaluate the properties of FVFT, particularly in terms of its ability to differentiate demented from cognitively intact elderly persons. Despite the finding that all VF scores produced good differentiation between the two groups, their validity and reliability were markedly compromised by their moderate correlation with the CMMSE and the disappointingly high temporal variability in test-retest performance. Nevertheless, these shortcomings were shown to be reasonably improved by the use of the composite score in the present study; this produced a far better outcome in terms of test-retest reliability and concurrent validity with the CMMSE, while preserving satisfactory sensitivity and specificity. Thus, it is recommended that the composite score be used in future assessment whenever possible in order to attain a more credible and reliable screening (e.g., Monsch et al., 1992).

In terms of demographic variables, our findings are consistent with previous studies showing that age is inversely correlated with the VF scores (Tomer & Levin, 1993). However, we could not replicate the finding that VF scores correlate with educational levels (Lezak, 1995). It is possible that the lack of correlation is due to the large proportion of illiterate subjects in our sample. Moreover, the finding that females have higher VF scores than males in some previous studies was not obtained in this study (Lezak, 1995). This is probably a result of different sampling characteristics, as the number of male subjects in our study is disproportionately smaller than that of female subjects.

In regard to the VF mood-related categories, the original aim for their inclusion in the FVFT was to evaluate the possible confound of depression or “pseudodementia” in the course of screening for an organic dementing process. This is based on the recent finding from cognitive research that depressed subjects tend to form associations or recall events with negative affective tone much more readily than events with positive affective tone (Dunbar & Lishman, 1984). As no difference in depression levels was found between the dementia and control groups, as shown by the CHDRS, no such mood dependent effect was, in fact, expected from the present study. In keeping with Fuld’s original finding (Fuld, 1980), however, a higher mean response was obtained across subjects for the “things that make you happy” category as compared with the “things that make you sad” category. Interestingly, in the present study, such tendency to report positive events was found to be more pronounced among the control subjects as compared to the dementia subjects. Although the reason for this group interaction effect remains somewhat elusive, it might relate to the difference in positive-versus-negative mood states between the two groups, which unfortunately has not been specifically addressed in the present study.

In the development of tests of cognitive function in cross-cultural research, there are two kinds of potential bias that have been eloquently discussed by Richards and Brayne (1996). The first refers to test bias arising mainly from inappropriate item selection, such as items that are poorly translated to another language or those involving objects that are unfamiliar to different cultures. The second concerns the patterns of experience that different cultures bring to the test situation, including factors such as education and occupation. These considerations are particularly pertinent in studies involving Chinese elderly.

For instance, it is our experience that tests on verbal fluency appear to be much more well received by elderly Chinese subjects, as they are not perceived to be as threatening as the other existing instruments, such as the CMMSE, which appears to be a more deliberate test of cognitive abilities. This is important, as we have found that Chinese elderly in general are rather reserved in displaying intellectual skills in front of unfamiliar people and in formal situations. In this respect, the VF tests are considered to be more “user-friendly,” as they only involve everyday items that elderly people are quite used to talking about.

Psychometrically speaking, however, other cognitive instruments, such as the CMMSE, have been found to possess better properties, particularly in terms of sensitivity and specificity (Chiu et al., 1994). This outcome is probably due to the composite nature of such scales, which normally involve assessment of a more comprehensive range of cognitive functions. As VF tests have been generally found to reflect problems specific to semantic and executive functions (Hodges, 1994), there seems to be more room for variability in test performance among patients, depending on the nature of dementia conditions. It is noteworthy that research on such differential performance characteristics is currently being conducted by several groups in the field, and their findings would certainly help to enhance our understanding of the specific cognitive changes associated with different types of dementia (Butters et al., 1986; Hanley et al., 1990; Hodges et al., 1990; Monsch et al., 1994).

Finally, several methodological issues require further attention. First, the inclusion of both cortical and subcortical patients in the dementia group might have confounded effects pertaining to these conditions. However, a direct comparison of these conditions was made infeasible due to the preponderance of subjects with Alzheimer’s disease in the study. Nevertheless, a separate analysis on the Alz-

---

Table 1. Optimal Cut-off Scores and Corresponding Rates of Sensitivity and Specificity for the Individual and Combined Verbal Fluency Categories

<table>
<thead>
<tr>
<th>Categories</th>
<th>Cut-off Scores</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>9</td>
<td>83.9</td>
<td>84.9</td>
</tr>
<tr>
<td>Fruits</td>
<td>6</td>
<td>94.6</td>
<td>81.1</td>
</tr>
<tr>
<td>Vegetables</td>
<td>7</td>
<td>78.2</td>
<td>71.7</td>
</tr>
<tr>
<td>Combined</td>
<td>19</td>
<td>80.0</td>
<td>88.7</td>
</tr>
</tbody>
</table>

---

*CHDRS, no such mood dependent effect was, in fact, expected from the present study. In keeping with Fuld’s original finding (Fuld, 1980), however, a higher mean response was obtained across subjects for the “things that make you happy” category as compared with the “things that make you sad” category. Interestingly, in the present study, such tendency to report positive events was found to be more pronounced among the control subjects as compared to the dementia subjects. Although the reason for this group interaction effect remains somewhat elusive, it might relate to the difference in positive-versus-negative mood states between the two groups, which unfortunately has not been specifically addressed in the present study.*
heimer subjects alone produced a pattern of results very similar to what was presented for all demented subjects in this report. Future studies are needed to examine possible differences among subtypes of dementia in terms of performance on the FVFT. Second, the comparability of the FVFT with the other verbal fluency studies was limited by the test duration recommended in the original manual (i.e., 60 s for the animal category and 30 s for the other categories). It seems apparent that the application of a more commonly endorsed duration (e.g., 60 s) in future studies would significantly improve the test. Third, as the diagnosticity of a test for a clinical condition often depends on its base rate in a given population (Tversky & Kahneman, 1973; Wright & Ng, 1982), findings regarding sensitivity and specificity in the present study should be treated cautiously. In this context, the ability of the FVFT as a screening instrument to offer additional predictive value would be increasingly difficult as the base rate deviates from our optimal condition (i.e., base rate of this study was about 0.5 due to a similar number of subjects in the two groups).

In conclusion, our preliminary findings suggest that the individual category fluency tests in the FVFT are characterized by unsatisfactory validity and reliability. On the other hand, the use of composite score in this study improves on these aspects significantly, and therefore is recommended for use in future investigation. To affirm the usefulness of the FVFT as a screening instrument for dementia in the community, further epidemiological studies on a larger scale are necessary.

ACKNOWLEDGMENT

Address correspondence to Dr. Helen Chiu, Department of Psychiatry, Prince of Wales Hospital, Shatin, Hong Kong.

REFERENCES


Received September 17, 1996
Accepted April 17, 1997