Comparison of EXIT-25 and the Frontal Assessment Battery for Evaluation of Executive Dysfunction in Patients Attending a Memory Clinic

Paige Moorhouse  Mary Gorman  Kenneth Rockwood
Dalhousie University, Halifax, N.S., Canada

Key Words
Dementia · Executive function · Frontal Assessment Battery · Cognition · Feasibility · EXIT-25

Abstract

Background/Aims: The Frontal Assessment Battery (FAB) and the EXIT-25 have emerged as 2 widely used screening instruments for executive dysfunction, but their screening properties have not been evaluated in a head-to-head comparison. Methods: Prospective cohort study of 92 individuals presenting for cognitive assessment at a community hospital. Results: The EXIT-25 took longer than the FAB to complete (mean difference = 9.27 min, 95% CI: 9.86–8.68). EXIT-25 and FAB scores showed significant correlation (Spearman’s r = –0.79, p < 0.001) with one another, and each showed acceptable convergent validity and divergent validity. Conclusions: The 2 tests provide similar information about the presence of executive dysfunction. The FAB takes less time to complete, and appears to be less frustrating for patients, making it more feasible as a screening test for executive dysfunction in a memory clinic setting.

Introduction

‘Executive function’ refers to the planning and performance of goal-directed behavior that requires social judgment, empathy, and the ability to anticipate consequences of one’s actions. Executive function can be broken into its principal components, which include volition, planning, purposive action, and effective performance [1, 2], with disorders of these functions classically localized to the frontal lobes, especially the prefrontal cortex [3, 4]. Impaired frontal lobe function is clinically important, as it can result in socially embarrassing conduct, social withdrawal, lack of initiative, and increasing reliance on routine, from which deviation becomes upsetting. Although not exclusive to dementia [5], the importance of executive dysfunction is increasingly recognized in the dementia syndrome [6–9]. Failure of executive function correlates with functional loss [3, 10] and with an increased risk of death [11]. Pragmatically, failure of executive function can compromise treatment planning and adherence [12].

Despite the importance of executive dysfunction to clinicians, there is as yet no accepted standard for diagnosing executive dysfunction beyond clinical diagnosis by an expert. Although neuroimaging techniques have been evaluated [13], the clinical evaluation of executive function appears to be most robust [14]. Even so, many cognitive screening instruments omit formal assessment of executive function [15]. To address this important omission, 2 instruments have emerged as widely used screening tests of executive function in clinic settings.

The Frontal Assessment Battery (FAB) was devised as a short bedside test of executive function [16]. The FAB includes a subset of 6 items to evaluate conceptualization, mental flexibility, motor programming, sensitivity to in-
terference, inhibitory control, and environmental autonomy. It has been found to be a feasible and sensitive assessment of frontal lobe dysfunction, and has been applied to several different patient populations. [17]. Reports of its ability to differentiate between Alzheimer’s and other types of dementia are, however, conflicting [18–20].

The EXIT-25 includes 25 items assessing perseveration, intrusions, apathy, disinhibition, utilization and imitation behaviors, concentration, motor impersistence, and frontal release signs. It is reported to take 10 min to administer and has shown good inter-rater reliability, correlations with other measures of executive function, and validity in institutionalized and noninstitutionalized settings [5]. To date, there are no published studies describing the screening properties of these 2 tests in a head-to-head comparison. Here, we aim to compare the feasibility of the FAB and the EXIT-25 in a busy clinic setting and to compare the validity of the 2 tests.

**Methods**

**Subjects**

The study was conducted at St. Martha’s Hospital, a university-affiliated geriatric medicine memory clinic at a regional hospital in Nova Scotia, Canada. There, both the Frontal Assessment Battery and the EXIT-25 tests are administered as part of routine cognitive assessment. The tests were administered by a geriatrician and by 2 advance practice nurses. In addition, for the assessment of content validation and evaluation of feasibility, we surveyed 3 other geriatricians based at the QEII Health Sciences Centre, a tertiary care teaching hospital, who were familiar with both tests.

The sample size was powered to detect an effect size of 0.7 (Hedge’s G, a variation of Cohen’s D that adjusts for bias due to small sample sizes) or 80% power to detect a 20% difference in time to complete, which gave a necessary sample size of 60.

Test scores were collected from 92 sequential patients without aphasia, deafness, or blindness, who presented for cognitive assessment from August 2005 to March 2006. Likert scale surveys were administered to the assessors and other geriatricians at the end of the study period to assess acceptability of the tests [21].

**Clinical and Cognitive Assessment**

Each patient underwent routine cognitive assessment including completion of a Comprehensive Geriatric Assessment form [22], the Mini-Mental Status Examination (MMSE) [23], the Physical Self-Maintenance Scale and Instrumental Activities of Daily Living Scale [24], the Functional Assessment Staging Tool [25], and the Brief Cognitive Rating Scale [26]. The Brief Cognitive Rating Scale includes tests of attention and concentration, constructional praxis and a semi-structured history of judgment, behavior, and initiation that tap aspects of frontal lobe function. In addition to the FAB and EXIT-25, the clock drawing test [27, 28] was also administered. The clock drawing test is a task to assess visuospatial and executive function in which the patient is asked to draw a clock and place the hands at a certain time (10 min after 11, or 20 min after 8). Various scoring methods are available. We used the scoring system of Stähelin et al. [27] in which a maximum of 6 points are available (for including the correct 12 numbers in the correct position and placing 2 distinguishable hands at the correct time). Abnormalities in the drawing can be easily assessed and compared across time to assess the course of the disease.

**Measures of Feasibility**

Because both the FAB and the EXIT-25 were already part of routine assessment in the clinic, assessment of feasibility consisted of recording the time to complete each test. Acceptability was measured using a Likert scale survey completed by all 5 test administrators at the end of the study period [21].

**Validation Strategy**

Both tests have been validated as part of their original development [4, 15]. Additional content validation was carried out by presenting each instrument to the raters and additional geriatricians, and eliciting their comments about validity and feasibility. Because there is no accepted gold standard upon which to base comparison of criterion validity between the tests, construct validity was assessed by correlating the EXIT-25 and FAB scores, first with each other using the Spearman correlation, and then with other instruments used in the assessment of dementia using k values and visual display of data. The scores were also compared with a clinical judgment of dementia, made using DSM-IV criteria [29]. Note that the physicians who made the diagnoses were not blind to the FAB and EXIT-25 test scores.

**Analysis**

Test scores were analyzed by an investigator who did not have access to identifying data. Feasibility was analyzed by a t test that compared the time taken (in minutes) to complete each test. Acceptability was analyzed by comparing the proportions who rated each Likert scale item as ‘agree’ or greater. The FAB and EXIT-25 scores were correlated with each other and against the other measures using Spearman’s rho. The proportion assigned as abnormal using previously published cutoffs for abnormal scores (FAB score <12 and EXIT-25 score ≥11) [15, 30] were compared using χ² tests.

**Results**

The baseline characteristics of the patients are shown in table 1. The FAB took significantly less time to complete (4.95 min, 95% CI: 4.62–5.29) compared with the EXIT-25 (14.23 min, 95% CI: 13.52–14.94) with a mean difference of 9.27 min (95% CI: 8.68–9.86, p < 0.01). No patient refused to complete the FAB, and 1 patient refused to complete the EXIT-25. Likert scale acceptability data indicates that 100% of testers (n = 5) agreed or strongly agreed that the FAB was easy to administer, compared with 60% who agreed or strongly agreed that the EXIT-25 was easy to administer. No testers indicated...
that the FAB caused frustration in patients compared with 80% of testers who indicated that the EXIT-25 caused frustration in patients. No testers reported patients asking for clarification of the test instructions during administration of the FAB, while 80% of testers reported patient requests for clarification during administration of the EXIT-25.

FAB and EXIT-25 scores showed a strong correlation with one another (Spearman $r = -0.79$, $p < 0.001$). The presence of an abnormal score on one test correlated with the presence of an abnormal score on the other test ($r = 0.42$, $p < 0.01$). However, no clear cutoff score for either test was associated with a final diagnosis of dementia of any type (Appendix). Of note, both tests include a test of verbal fluency, being the number of words (excluding proper nouns) beginning with the letter ‘S’ (FAB) or ‘A’ (EXIT-25) that are generated within 1 min. Patients generated 6.3 (95% CI: 5.0–7.5) words on the FAB, compared with 3.8 words (95% CI: 2.8–4.8) on the EXIT-25.

Both tests showed modest correlations with the MMSE score (table 2). There was no significant difference between tests in terms of their correlation with other measures of executive function. Both the FAB and EXIT-25 showed modest correlations with the CDT score (0.59 and –0.61, respectively) and the IADL score (–0.46 and 0.40, respectively). Neither test score correlated with the number of medications or medical problems of the individual indicating appropriate divergent validity.

### Discussion

Given the importance of screening for executive dysfunction in patients referred for memory disorders, we evaluated the properties of 2 commonly used screening tests in a regional referral hospital memory clinic. Each showed similar construct validity, but the FAB was favored for ease of use and acceptability to patients.

Our data must be interpreted with caution. Although the sample was well powered for the analyses, it is still not large. Even so, by evaluating consecutive referrals, we have a good sense of performance under real-world conditions. This too is strengthened by producing estimates from a regional referral hospital, which is less likely to have as strong a referral filter as a tertiary care centre, and therefore more generalizable estimates. Our study included no criterion reference for executive function, reflecting the absence of a ‘gold standard’ for such diagnoses. We were however able to correlate the 2 tests with other cognitive measures, including others that test executive functioning. Although we compared the results with a clinical diagnosis of dementia, the clinicians who made the dementia diagnosis were not blind to the test results. Even so, they knew the results of both sets of tests, and also had other grounds upon which to make a dementia diagnosis. We note that there remain questions about the ecological validity of going from clinical assessments, especially caregiver observations, about impaired executive function and standard ‘frontal lobe’ tests used clinically. Finally, we recognize that this study does not assess responsiveness of either instrument, or their predictive validity. It does, however, set the stage for these and additional studies to be undertaken.

The significant correlation of both tests with the MMSE deserves mention. The EXIT-25 demonstrated a similar degree of correlation in the original publication [4], in which this was described as a strength of the EXIT-25. However, Dubois et al. [16] did not find a correlation between FAB and MMSE scores in their original study, and also described this as a strength of the FAB. In a memory clinic setting, it is easy to imagine how patients who were impaired enough to have lower MMSE scores

---

**Table 1.** Baseline characteristics of the sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female ratio</td>
<td>0.61</td>
</tr>
<tr>
<td>Age, years</td>
<td>77.2 ± 8.9</td>
</tr>
<tr>
<td>Education, years</td>
<td>10 ± 3.5</td>
</tr>
<tr>
<td>MMSE score</td>
<td>21.5 ± 5.1</td>
</tr>
<tr>
<td>EXIT-25 score</td>
<td>16.5 ± 8.7</td>
</tr>
<tr>
<td>FAB score</td>
<td>10.6 ± 4.2</td>
</tr>
</tbody>
</table>

Data presented as means ± SD.

**Table 2.** Spearman’s correlations of the FAB and the EXIT-25 interview scores with other measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>FAB, r (p)</th>
<th>EXIT-25, r (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE score</td>
<td>–0.58 (p &lt; 0.001)</td>
<td>0.61 (p &lt; 0.001)</td>
</tr>
<tr>
<td>IADL score</td>
<td>–0.46 (p &lt; 0.001)</td>
<td>0.40 (p &lt; 0.001)</td>
</tr>
<tr>
<td>Clock score</td>
<td>0.59 (p &lt; 0.001)</td>
<td>–0.61 (p &lt; 0.001)</td>
</tr>
<tr>
<td>ADL/IADL score from CGA</td>
<td>–0.56 (p &lt; 0.001)</td>
<td>0.53 (p &lt; 0.001)</td>
</tr>
<tr>
<td>Number of medications listed on CGA</td>
<td>–0.13 (p = 0.22)</td>
<td>0.11 (p = 0.29)</td>
</tr>
<tr>
<td>Number of medical problems listed on CGA</td>
<td>0.028 (p = 0.79)</td>
<td>–0.044 (p = 0.69)</td>
</tr>
</tbody>
</table>

(I)ADL = (Instrumental) activities of daily living; CGA = Comprehensive Geriatric Assessment.
would also show an important degree of executive dysfunction. In a setting in which fewer people had dementia, the correlation might not be as strong.

Evaluation of some aspects of executive function can be challenging in a clinical setting. For example, tests of sustained attention have been shown to correlate with volition, but there are currently no cognitive tests for volition per se [31]. Similarly, effective performance (the ability to self-monitor for errors, and regulate behavior) may be reflected in the patient’s insight towards cognitive deficits [7]. Purposeful action involves initiation, maintenance, and modulation of behaviors culminating in the delivery of tasks. Claims have been made for its assessment as part of tests of fluency, attention, concept formation, and sorting [2]. This can seem at a distance, however, from the problem of decreased intention to undertake common tasks, which is important part of how caregivers understand dementia. Indeed, caregivers may identify the failure to undertake tasks as the person’s ‘not being themselves’ [32]. Failure to assess executive function in a way that is comprehensible to caregivers can put a gap between what we seem to assess and what carers seem to care about. Especially in the context of a busy clinic, feasibility is an important consideration. As we move to a greater appreciation of the determining importance of executive function on what we recognize clinically as dementia [5], it is important that an efficient measure of dementia be available for widespread use.

### Appendix: Questions Pertaining to Acceptability of FAB and EXIT-25

I found the test to be easy to administer:

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>No opinion</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
</table>

I found the test to be easy to administer the same way each time:

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>No opinion</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
</table>

I found the test caused frustration in patients:

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>No opinion</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
</table>

I found that patients asked for clarification of the items on the test:

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>No opinion</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
</table>

Since the time that data collection began has anyone refused to complete the test? Yes ☐ No ☐

If yes, please estimate how many: __________

### Acknowledgments

The authors would like to thank Patty McIsaac and Betty Ann Arbuckle for their assistance with data gathering.

### References


