Reliability, Validity, and Clinical Utility of the Executive Function Performance Test: A Measure of Executive Function in a Sample of People With Stroke

Carolyn M. Baum, Lisa Tabor Connor, Tracy Morrison, Michelle Hahn, Alexander W. Dromerick, Dorothy F. Edwards

This study examined the reliability and validity of the Executive Function Performance Test (EFPT). The EFPT assesses executive function deficits in the performance of real-world tasks. It uses a structured cueing and scoring system to assess higher-level cognitive functions, specifically initiation, organization, sequencing, safety and judgment, and task completion. Seventy-three participants with mild to moderate stroke and 22 age- and education-matched controls completed the 4 EFPT tasks (cooking, using the telephone, managing medications, and paying bills). Significant differences were found between participants with mild and moderate stroke and healthy control participants. The EFPT can help occupational therapists determine the level of support needed by people with cognitive impairments to perform complex instrumental tasks. Objective information derived from this assessment is an essential part of the process of determining whether the person can live independently and helping families understand how to support the performance of their family members at home.


Executive abilities, or higher-level cognitive functions, enable people to successfully formulate goals, plan how to achieve them, and carry out the plans effectively (Kaye, Grigsbys, Robbins, & Korzun, 1990; Lezak, 1982, 2004; Stuss, 1992). The ability to make decisions, self-correct, and use judgment enables the performance of complex activities of daily living (ADLs) essential for functional independence (Burgess, 2000; Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Burgess et al., 2006; Crawford & Channon, 2002; Goel, Grafman, Tajik, Gana, & Danto, 1997; Lezak, 1982, 2004). These higher-order cognitive abilities underlie and support daily life performance; their impairment or loss compromises a person’s ability to fully participate in society (Goel et al., 1997; Katz & Hartman-Maeir, 2005; Lezak, 1982).

Executive functions traditionally have been assessed through standardized psychometric measurements administered in controlled environments (Lezak, 1995). Although neuropsychological measures provide good indicators of fundamental cognitive and executive components, performance on neuropsychological tests often is not predictive of real-world complex task performance and functional ability (Alderman, Burgess, Knight, & Henman, 2003; Burgess et al., 2006; Gioia & Isquith, 2004; Keil & Kaszniak, 2002; Shallice & Burgess, 1991; Wilson, 1993). Daily life performance and the executive abilities that support it often require multitasking and the generation and implementation of adaptive strategies to accommodate to novel environments and perform tasks in the real world (Manchester, Priestley, & Jackson, 2004). Assessment of the full array of executive functions...
necessary for complex life tasks requires testing be conducted in real-world environments (Burgess et al., 2006; Goverover, 2004). Although neuropsychologists acknowledge that real-world tasks are sensitive to brain damage, they have indicated that the testing process is “too unwieldy” to be a part of a routine neuropsychological evaluation (Godbout, Grenier, Braun, & Gagnon, 2005). Occupational therapists may instead provide this information if they have valid tools to measure real-world executive performance.

Traditionally, occupational therapists have been asked to determine a person’s capacity to be safe, live independently, be gainfully and purposefully employed, and participate in meaningful activities. To address those issues, occupational therapists assess everyday task performance to determine strengths, limitations, and challenges that the person with cognitive impairment will face in performing ADLs (Baum & Edwards, 1993). As we become able to answer questions about people’s occupational performance needs, our team has elected to study everyday life issues from the perspective of the Person–Environment–Occupational Performance Model (Baum & Christiansen, 2005), which looks at both the intrinsic and the extrinsic factors that enable occupational performance and participation. Each of these factors must be studied in the context of everyday life performance. The cognitive abilities of particular interest in evaluating occupational performance include initiation, the process that precedes the performance of a task (DePoy, Maley, & Stranraugh, 1990; Kaye et al., 1990; Lezak, 2004; Weld & Evans, 1990); organization, the physical arrangement of the environment, tools, and materials to facilitate efficient sequencing or effective performance (Lezak, 2004; Weld & Evans, 1990); judgment (Goel et al., 1997; Lezak, 1982); and task completion (Goel et al., 1997).

As crucial as executive processes are for daily function, few assessments have been specifically designed to capture them in everyday life (Burgess et al., 2006; Goel et al., 1997; Lezak, 1982). Recently, interest has increased in the development of real-world structured tasks designed to capture executive deficits through observation of single or multiple complex tasks (Baum & Edwards, 1993; Bechara, Damasio, Damasio, & Anderson, 1994; Gaudette & Anderson, 2002; Goel et al., 1997; Shallice & Burgess, 1991). Real-world performance tests are ecologically valid because they use naturalistic environments, do not provide artificial structure, require multitasking, and reflect the press of everyday task performances (Baum & Edwards, 1993; Bechara et al., 1994; Burgess et al., 2006; Gaudette & Anderson, 2002; Goel et al., 1997; Shallice & Burgess, 1991). The development of these practical tests is critical to the field of occupational therapy and to cognitive rehabilitation because clinicians must assess their clients’ potential to return home and regain functional independence (Baum & Edwards, 1993; Gaudette & Anderson, 2002).

Occupational therapists have developed several assessments to measure performance of instrumental tasks, including the Allen Cognitive Levels Test Battery (Allen, Earhart, & Blue, 1992) and the Assessment of Motor and Process Skills (Fisher, 1993). Although these validated measures indicate problems a person is experiencing in cognitive and process skills during performance of an everyday task (Baum & Edwards, 1993; Fisher, 1993; Linden, Boschian, Eker, Schalen, & Nordstrom, 2005; Mercier, Audet, Hebert, Rochette, & Dubois, 2001; Nygard, Bernspand, Fisher, & Winblad, 1994; Park, Fisher, & Velozo, 1994; Secrest, Wood, & Tapp, 2000; Velligan, True, Lefton, Moore, & Flores, 1995), they do not record the person’s capabilities when provided with progressive levels of support. A third measure, the Kitchen Task Assessment (Baum & Edwards, 1993), which assesses the capacity to perform a simple cooking task, evaluates the executive functions of initiation, organization, sequencing, safety and judgment, and completion, and it records capacities exhibited with progressive support. Its primary limitation is that it assesses performance on only a single task.

The measure described in this article, the Executive Function Performance Test (EFPT; Baum, Morrison, Hahn, & Edwards, 2003; Katz, Tadmore, Felzen, & Hartman-Maer, 2007), has several advantages over existing performance-based assessments. First, it is easily administered after brief training (provided in the manual) in what to observe and how to cue and score. Second, it isolates cognitive components related to executive functions during performance of four instrumental activities; this information is crucial to the generation of treatment plans. Third, it uses a top-down approach that allows the practitioner to objectively assess the client during the performance of a task, and unlike many other instruments assessing instrumental activities of daily living (IADLs), it assesses actual performance rather than rely on proxy or self-report. Moreover, the EFPT measures the level of support the person needs from another person to be successful in the four daily life tasks central to community living: (1) preparing or heating up a light meal, (2) managing medications, (3) using the telephone, and (4) paying bills (Lysack, Neufeld, Mast, Macneill, & Lichtenberg, 2003).

Development of the EFPT Measure

The EFPT was developed as a measure of executive performance. It records what the person can do and the level of support necessary to successfully perform a task. It is designed to provide the practitioner with information that will help family members understand and support their loved ones’ performance. The EFPT also was designed to be sensitive to
degrees of performance breakdown accompanied by various levels of disease progression or severity.

The EFPT uses the format of the Kitchen Task Assessment (Baum & Edwards, 1993). The EFPT replaces the task of preparing cooked pudding with preparing cooked oatmeal (the two tasks require the same number of steps), and it is further expanded to include the assessment of the support needed to complete a telephone call, manage medication, and pay a bill. In addition, the EFPT enhances the method of recording cues to support performance. Unlike other tests of cognitive function (with the exception of the Kitchen Task Assessment), the EFPT does not assess what people cannot do; rather, it identifies what they can successfully accomplish. The EFPT’s standardized cueing system makes it possible to identify a wider range of abilities in people who may not be able to perform the task if they were measured without support. This cueing system is based on the progressive need for assistance associated with increasing levels of cognitive impairment and gives the tester a straightforward way to record the assistance required to successfully perform the task. This information may provide the basis for using learning strategies to help family members or care providers gain skills to enable the performance of the person with cognitive impairment.

The EFPT’s reliability and validity have recently been established in two clinical populations. Goverover et al. (2005) reported its validity with a sample of adults with multiple sclerosis, and Katz, Tadmore, Felzen, and Hartman-Maier (2007) reported its reliability and validity in people with schizophrenia. In the present study, we sought to establish the EFPT’s reliability and validity in a population of people with stroke and to demonstrate its sensitivity to stroke severity. Specifically, we sought to (1) determine the EFPT’s interrater reliability and internal consistency, (2) examine its construct validity in people with mild and moderate stroke and a sample of healthy control participants, and (3) evaluate its concurrent validity using a battery of neuropsychological tests. We hypothesized that healthy control participants would perform significantly better on the EFPT than people with stroke. We further hypothesized that participants with mild stroke would perform significantly better than those with moderate stroke. Finally, we hypothesized that EFPT scores would show significantly moderate correlations with neuropsychological assessments of executive function.

Method

Participants

Study participants were recruited through the Cognitive Rehabilitation Research Group of Washington University in St. Louis. The Cognitive Rehabilitation Research Group registers participants from the acute Neurology Stroke service of Barnes–Jewish Hospital; all data were obtained in compliance with and with approval from the Washington University Institutional Review Board. A nurse coordinator prospectively evaluated and recorded clinical and demographic information for all participants, and a neurologist confirmed the diagnosis of stroke. Permission to contact participants for additional studies was obtained at the time of the acute stroke admission. Participants in this study were admitted from July 2002 to June 2004. Six months after stroke onset, participants were contacted and asked to enroll in this study. All participants were fully independent before their stroke as indicated by prestroke Barthel Index (Mahoney & Barthel, 1965) scores of ≥90 and Modified Rankin Index scores of ≥2. All study participants resided in community settings at the time of the study assessment. Healthy control participants were recruited through the Volunteers for Health Program at Washington University. Control participants were screened for cognitive impairment and health conditions known to affect cognitive performance before study enrollment. Informed consent was obtained from all study participants.

We tested 73 participants with stroke approximately 6 months after stroke onset. None of the participants was in a rehabilitation program. The 6-month time frame was chosen to allow for natural recovery to occur. Participants with stroke were divided into two groups on the basis of their admission score on the National Institutes of Health Stroke Scale (Brott et al., 1989). Scores of 5 or less are considered mild, and scores between 6 and 15 represent a moderate level of stroke severity (Brott et al., 1989; Edwards, Hahn, Baum, & Dromerick, 2006). We also assessed 22 healthy control participants. The characteristics of study participants are presented in Table 1.

Measures Used in the Study

The specifics of administering each of the four instrumental tasks are described in the test manual and can be obtained from the Cognitive Rehabilitation Research Group’s Web site, http://crg.wustl.edu/outcome_assessment.html. The “simple cooking” task requires the person to prepare quick-cooking oatmeal following written instructions on the package (or on a sheet prepared with large print). “Using the telephone” includes looking up a grocery store number in the telephone book, calling the store, and asking whether the store delivers groceries. “Managing medications” requires the person to select the prescription medication from three available distracter drugs and take it with food as required by the directions (the pill is a sugar-free placebo). To evaluate “paying bills,” two bills, checks, a check register, and mail to serve
as a distracter are provided in an envelope. The person is required to locate the two bills in the stack of mail, pay them according to the money available in the account, and balance the account.

Before beginning each task, the person is asked about familiarity with the task and whether he or she performs it independently or with assistance. All necessary materials are provided for the assessment in a box. In all four tasks, the EFPT assesses the person’s ability to use five executive functions (components) of a task: (1) initiation of a task (beginning the task), (2) organization (retrieval and arrangement of tools), (3) sequencing (execution of steps in a correct order), (4) safety and judgment (avoids a dangerous situation), and (5) completion (deciding and acknowledging when a task is complete). Five levels of cueing can be delivered: 0 (no cue required); 1 (verbal guidance); 2 (gestural guidance); 3 (direct verbal assistance); 4 (physical assistance); and 5 (do for the participant). A higher score reflects a need for more cueing and indicates more severe executive function deficits. People with motor impairment are scored according to the cue level they need but are not penalized if they ask for help because the impairment necessitates physical assistance.

The highest level of cueing necessary to support task performance is recorded; thus, the test results in three scores: (1) the executive function (EF) component score, (2) the task score, and (3) a total score. The EF component score is calculated by summing the numbers recorded on each of the four tasks for initiation, organization, sequencing, safety and judgment, and completion. Scores on each EF component can range from 0 to 5, and the total for all four tasks can range from 0 to 20. The task score is calculated by summing the five scores for each task. The range for each task is 0 to 25. The total score is the sum of the performance on all four tasks; the total score of performance on all four tasks can range from 0 to 100. The forms for the EFPT can be viewed at http://crrg.wustl.edu/outcome_assessment.html.

**Measures Used for Classification and Exclusion**

**National Institutes of Health Stroke Scale.** The NIHSS (Brott et al., 1989) assesses cognitive, sensory, and motor impairments resulting from a stroke. The 13-item test is based on a total score ranging from 0 to 46; a lower score indicates a lower level of impairment. We used the test to define mild and moderate stroke.

**Short Blessed Test (Memory, Orientation, and Concentration).** We used the Short Blessed Test (Katzman et al., 1983) to measure cognition. The test consists of six items assessing memory, orientation, and concentration and is a reliable and valid screening tool for detection of dementia in community

---

Table 1. Demographic Characteristics and Performance on Executive Function Performance Test (EFPT) by Group

<table>
<thead>
<tr>
<th>Demographic Characteristic or Test</th>
<th>Control (N = 22)</th>
<th>Mild Stroke (N = 59)</th>
<th>Moderate Stroke (N = 14)</th>
<th>χ²(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n %</td>
<td>n %</td>
<td>n %</td>
<td></td>
</tr>
<tr>
<td>Race**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>6 (27)</td>
<td>22 (37)</td>
<td>4 (29)</td>
<td>0.91</td>
</tr>
<tr>
<td>White</td>
<td>16 (73)</td>
<td>37 (63)</td>
<td>10 (71)</td>
<td>0.75</td>
</tr>
<tr>
<td>Gender†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6 (27)</td>
<td>15 (25)</td>
<td>3 (21)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>16 (73)</td>
<td>44 (75)</td>
<td>11(79)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>59.45 (15.78)</td>
<td>64.57 (14.28)</td>
<td>64.14 (14.78)</td>
<td>1.10</td>
</tr>
<tr>
<td>Education (years)</td>
<td>14.68 (2.23)</td>
<td>12.81 (2.90)</td>
<td>12.5 (3.25)</td>
<td>1.23</td>
</tr>
<tr>
<td>NIHSS</td>
<td></td>
<td>2.00 (1.46)</td>
<td>10.64 (2.99)</td>
<td></td>
</tr>
<tr>
<td>EFPT total score***</td>
<td>1.51 (2.27)</td>
<td>7.87 (8.42)</td>
<td>24.21 (5.39)</td>
<td>15.49</td>
</tr>
<tr>
<td>EFPT tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooking**</td>
<td>1.23 (2.81)</td>
<td>2.98 (4.90)</td>
<td>5.57 (7.27)</td>
<td>9.28</td>
</tr>
<tr>
<td>Using telephone***</td>
<td>0.09 (0.29)</td>
<td>1.83 (2.27)</td>
<td>6.57 (9.00)</td>
<td>18.00</td>
</tr>
<tr>
<td>Medications***</td>
<td>0.42 (1.40)</td>
<td>0.92 (1.34)</td>
<td>5.50 (8.99)</td>
<td>13.52</td>
</tr>
<tr>
<td>Paying bills**</td>
<td>0.23 (0.69)</td>
<td>1.92 (2.19)</td>
<td>4.43 (6.94)</td>
<td>11.07</td>
</tr>
<tr>
<td>EFPT EF components</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiation</td>
<td>0.00 (0.00)</td>
<td>0.83 (1.67)</td>
<td>3.43 (7.14)</td>
<td>8.47</td>
</tr>
<tr>
<td>Organization***</td>
<td>0.13 (0.34)</td>
<td>1.77 (2.22)</td>
<td>6.21 (8.39)</td>
<td>18.97</td>
</tr>
<tr>
<td>Sequencing***</td>
<td>0.58 (1.05)</td>
<td>3.08 (3.20)</td>
<td>6.93 (7.87)</td>
<td>18.33</td>
</tr>
<tr>
<td>Safety and judgment**</td>
<td>0.14 (0.47)</td>
<td>1.32 (2.19)</td>
<td>4.07 (6.67)</td>
<td>7.76</td>
</tr>
<tr>
<td>Completion**</td>
<td>0.11 (0.49)</td>
<td>0.88 (1.90)</td>
<td>3.57 (7.08)</td>
<td>8.76</td>
</tr>
</tbody>
</table>

**Note.** NIHSS = National Institutes of Health Stroke Scale; EF = executive function.

**1p < .63.** **2p < .68.** **3p < .005.** **4p < .0001.**
and long-term-care populations. Possible scores range from 0 to 28. The higher the score is, the greater the cognitive impairment is. We did not test people with scores >12; thus, people with dementia were excluded.

Neuropsychological Tests

Animal Naming. The Animal Naming test (Barr & Brandt, 1996) is a measure of verbal fluency that asks the participant to generate as many animal names as possible in 60 s. It assesses the efficiency of verbal retrieval, short-term memory, and cognitive flexibility.

Trailmaking Test. The Trailmaking Test (Reitan & Wolfson, 1995) provides information regarding attention, visual scanning, and executive function. Part A requires the participant to draw lines and connect 25 numbers scattered on a page. Part B calls for the connection of numbers and letters in order, alternating between letters and numbers. Two scores reflect the total time in seconds required to complete each task. An oral version is available for people with motor deficits.

Wechsler Memory Scale–Revised

Digit Span Forward and Backward Test. The Wechsler Memory Scale–Revised, Digit Span Forward and Backward Test (Wechsler, 1987), measures attention and verbal working memory. The examiner is required to verbally present digits at a rate of one per second. The forward test requires the participant to repeat the digits verbatim. The backward test requires the participant to repeat the digits in reverse order.

Logical Memory Total Recall Test. The Logical Memory Total Recall Test, also by Wechsler (1987), examines the ability to recall ideas in two orally presented stories. The test assesses attention, concentration, and verbal memory. Higher scores are indicative of better memory performance.

Functional Outcome Measures

FIM. The interview version of the FIM (Chang, Slaughter, Cartwright, & Chan, 1997) uses 18 items to grade the level of cognitive and physical assistance necessary for function. Item scores range from 1 (completely dependent) to 7 (completely independent). Two summary scores (mobility/self-care and communication/cognition) are derived. Total possible scores are 91 and 35, respectively. Higher scores represent greater independence.

Functional Assessment Measure. The Functional Assessment Measure (FAM; Hall, Hamilton, Gordon, & Zasler, 1993; Hawley, Taylor, Hellawell & Pentland, 1999) was designed to supplement the FIM; it has 12 items that further assess motor abilities, community mobility, communication, psychosocial adjustment, and cognition. The items are scored on the 7-point scale used in the FIM. Maximum total score for the scale is 84, and higher scores indicate greater functional independence.

Procedures

All participants were tested individually in a laboratory kitchen designed to simulate a home environment. The room was quiet and free from distractions. The EFPT tasks were administered in the same order (cooking, using the telephone, managing medications, paying bills). We obtained neuropsychological and functional measures during separate test sessions at 6 months. Demographic and stroke severity information for participants with stroke was obtained from the medical record for the acute hospital admission, and demographic data for control participants were obtained at the time of testing.

We conducted all data analyses using SPSS for Windows (Version 13: SPSS Inc., Chicago) and computed descriptive statistics for all variables. One-way analysis of variance (ANOVA) and chi-square analyses were used as appropriate to test for differences in demographic and neuropsychological variables across the groups. Intraclass correlation coefficients (ICCs) were computed to establish interrater reliability. Cronbach’s alpha was used to examine internal consistency.

We computed one-way ANOVAs to test the hypotheses of differences in performance across the three groups. An a priori adjustment of the criterion $p$ value for significance was computed by dividing .05 by 9 (the number of ANOVAs computed). This calculation resulted in an adjusted criterion of $p < .005$.

Given the number of measures and groups, post hoc analyses used Bonferroni tests corrected for multiple comparisons. Pearson correlation coefficients were used to determine criterion validity. The median scores for all EFPT variables were 0 for the control participants. Given this lack of variability, we used only the scores of the stroke participants for this analysis.

Results

The characteristics of the study sample are presented in Table 1. The mean age of the participants was 64.49 ($SD = 14.28$, range = 30–90 years). Although the control participants were slightly younger than the participants with stroke, no significant differences in age were found among the three groups. The control participants had more years of education than the stroke group (14.68 years vs. 12.81 years, respectively), although this difference was not statistically significant. The participants were predominantly female and White. No significant differences were found in race ($\chi^2 = 0.91, p = .63, df = 5$) or gender ($\chi^2 = 0.75, p = .68, df = 5$) across the
groups. The mean National Institutes of Health Stroke Scale scores were 2.00 (SD = 1.50) for the mild group and 10.64 (SD = 2.99) for the moderate group.

We computed a series of one-way ANOVAs to examine differences across the control, mild, and moderate stroke groups on the neuropsychological tests. In each case, the mean scores of the control group were better than those for the mild stroke group. The mild stroke group performed better than the moderate group on all variables. Significant differences were observed for all variables except Trails A. The means and standard deviations of these measures are shown by group in Table 2.

Reliability

We determined interrater reliability with three trained raters who simultaneously rated 10 participants, 5 with mild stroke and 5 healthy controls. The total score and each of four subtest scores were analyzed, and ICCs were computed. For the total EFPT score, the ICC was .91, and subtest ICC scores were .94 for the cooking task, .89 for paying bills, .87 for managing medication, and .79 for using the telephone. These coefficients are indicative of high levels of interrater reliability.

The internal consistency of the EFPT for the total sample was high (α = .94); subtest Cronbach’s alpha coefficients were .86 for the cooking task, .78 for paying bills, .88 for managing medication, and .77 for using the telephone. These coefficients support the internal consistency of the EFPT. Correlations between each of the test domains and the total score were as follows: initiation, r = .91; organization, r = .93; sequencing, r = .88; safety and judgment, r = .78, and completion of all steps, r = .89.

Construct Validity

The mean EFPT total score and the scores for each task and EF component are presented by group in Table 1. To examine the construct validity of the EFPT and EF components, we examined the scores across the three groups of participants. Construct validity is established if a test can discriminate between people with and without a known trait (Portney & Watkins, 2000). A series of one-way ANOVAs comparing performance across the three groups was computed separately for the total scores, tasks, and EF components; those findings are also presented in Table 1. As hypothesized, with the exception of initiation, we found significant differences among the groups for each analysis. Inspection of the mean EFPT total scores, task scores, and EF component scores indicated that the control participants had the lowest (best) scores, followed by the mild stroke group. The participants with moderate stroke had higher mean scores on all measures.

We then computed post hoc Bonferroni pairwise comparisons for each of the EFPT tasks and components. The results of the pairwise comparisons are presented in Table 2. The hypothesized differences between the groups were supported. The control EFPT total scores were significantly lower than the mild stroke scores (p < .05) and moderate stroke scores (p < .0001). Mild scores were also significantly lower than moderate scores (p < .0001).

Two of the subtests, Cooking and Paying Bills, significantly discriminated between control and mild stroke participants. The mild stroke participants performed less well than the healthy controls. Mild and moderate stroke groups were significantly different on three of the four subtests. Only the cooking subtest failed to significantly discriminate between the mild and moderate groups.

The test EF components were also examined. We found significant differences between control and mild stroke participants for sequencing (p < .001) and organization (p < .04). Differences between mild and moderate stroke groups were significant for organization (p < .0001), sequencing (p < .001), safety and judgment (p < .004), and completion (p < .01). These findings are presented in Table 3.

Criterion Validity

We examined concurrent validity, a form of criterion validity, by comparing scores on a battery of neuropsychological tests to performance on the EFPT. Concurrent validity is established when high correlations are found between the new measure and well-established measures of the phenomenon (Portney & Watkins, 2000). Only data from the stroke participants were used in this analysis. The results of these analyses are presented in Table 4. The hypothesis regarding the criterion validity of the EFPT was confirmed. Significant moderate correlations were found between the EFPT total score and neuropsychological measures assessing working memory, verbal fluency, and attention. These correlations include the recall score of the Wechsler Memory Scale (r = −.59), Animal Fluency (r = −.47), Trails B (r = .39), and

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control (M ± SD)</th>
<th>Mild Stroke (M ± SD)</th>
<th>Moderate Stroke (M ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trails A (seconds)***</td>
<td>31.0 ± 10.8</td>
<td>71.9 ± 63.6</td>
<td>188.3 ± 114.5</td>
</tr>
<tr>
<td>Trails B (seconds)**</td>
<td>73.8 ± 29.4</td>
<td>184.1 ± 98.5</td>
<td>279.6 ± 64.7</td>
</tr>
<tr>
<td>Digits Forward**</td>
<td>9.2 ± 2.6</td>
<td>6.5 ± 1.3</td>
<td>6.1 ± 2.3</td>
</tr>
<tr>
<td>Digits Backward**</td>
<td>5.3 ± 1.6</td>
<td>3.5 ± 1.7</td>
<td>3.1 ± 1.6</td>
</tr>
<tr>
<td>Story Recall***</td>
<td>30.6 ± 6.9</td>
<td>24.8 ± 8.1</td>
<td>28.0 ± 9.1</td>
</tr>
<tr>
<td>Letter Fluency**</td>
<td>38.3 ± 12.7</td>
<td>14.0 ± 7.5</td>
<td>25.3 ± 12.7</td>
</tr>
<tr>
<td>Animal Fluency***</td>
<td>22.6 ± 4.9</td>
<td>14.8 ± 5.5</td>
<td>8.8 ± 5.1</td>
</tr>
</tbody>
</table>

Note: Values are one-way analyses of variance comparing scores across groups. *p < .05. **p < .01. ***p < .001.
Table 3. Significant Between-Group Differences for Executive Function Performance Test (EFPT) Task and Component Scores by Group (p Values)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control vs. Mild</th>
<th>Control vs. Moderate</th>
<th>Mild vs. Moderate</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFPT total</td>
<td>.05</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>Subtest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooking</td>
<td>.008</td>
<td>.0001</td>
<td>.13</td>
</tr>
<tr>
<td>Using telephone</td>
<td>.06</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>Medications</td>
<td>.07</td>
<td>.0001</td>
<td>.001</td>
</tr>
<tr>
<td>Paying bills</td>
<td>.03</td>
<td>.0001</td>
<td>.01</td>
</tr>
<tr>
<td>Executive component</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiation</td>
<td>.34</td>
<td>.001</td>
<td>.004</td>
</tr>
<tr>
<td>Organization</td>
<td>.04</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>Sequencing</td>
<td>.001</td>
<td>.0001</td>
<td>.001</td>
</tr>
<tr>
<td>Safety and judgment</td>
<td>.77</td>
<td>.0001</td>
<td>.004</td>
</tr>
<tr>
<td>Completion</td>
<td>.47</td>
<td>.0001</td>
<td>.01</td>
</tr>
</tbody>
</table>

Digits Backward \( r = -.49 \). The EFPT total score was also significantly correlated with the Short Blessed Scale \( r = .39 \).

Lower, nonsignificant correlations were observed for tests that are not thought to assess executive function. These include Trails A \( r = .21 \) and Digits Forward \( r = -.26 \).

Differences in the correlations between the EFPT’s scores and the FIM and FAM scores provide additional evidence of criterion validity. The FIM is a measure of basic ADLs, whereas the FAM assesses the more complex instrumental skills needed for community independence. The correlations between the EFPT scores and the FIM \( r = -.40 \) and FAM \( r = -.68 \) scores reflect these differences. The higher correlation between the EFPT and FAM scores further supports the criterion validity of the EFPT as a measure important for the assessment of IADLs.

Discussion

This study examined the EFPT’s reliability and validity in healthy control participants and people with mild to moderate stroke. The need to more fully understand the effects of executive function deficits on complex task performance is well-documented (Kounti, Tsolaki, & Kiosseoglou, 2006). Deficits in working memory, attention, organization, and self-control have been linked to impaired performance of IADLs such as money management, taking medications, shopping, and driving (Grigson, Kaye, Baxter, Shetterly, & Hamman, 1989). However, no consensus exists on the most appropriate method for assessing executive function deficits when the primary interest is the prediction of performance of complex life tasks (Burgess et al., 2006). A growing body of literature suggests that, although neuropsychological measures are good indexes of isolated cognitive and executive components, they often are less effective in predicting real-world complex task performance and functional ability (Alderman et al., 2003; Burgess et al., 1998; Gioia & Isquith, 2004; Keil & Kasznik, 2002; Shallice & Burgess, 1996; Wilson, 1993).

Several recent articles have supported the assertion that assessment of executive functions necessary for complex life tasks is best conducted in real-world environments (Burgess et al., 2006; Godbout, Grenier, Braun, & Gagnon, 2005; Goverover et al., 2005). Such assessments will provide a better understanding of the impact of environmental factors on performance (Goverover et al., 2005). The EFPT is such a tool. It uses real rather than simulated activities as the assessment process, and the assessment is conducted in a real-world environment.

We found the EFPT to be a reliable and valid assessment of executive function abilities in people with mild to moderate stroke. Our data support the construct, criterion, and discriminant validity of the measure in this population. Similar findings have been reported with the EFPT for people with multiple sclerosis (Goverover et al., 2005) and schizophrenia (Katz et al., 2007). In addition to confirming the results of these investigations, our study provides additional evidence of criterion validity through the comparison of EFPT scores to accepted neuropsychological measures of executive function. Examination of the correlation coefficients between the EFPT and the neuropsychological tests illustrates the argument made by Burgess et al. (2006) in their recent call for the use of measures based on real-world performance in clinical and experimental studies of executive function. The moderate relationship between the neuropsychological tests and the EFPT suggests that, although both types of tests tap similar constructs, neuropsychological tests cannot fully capture the complexity of performance in context and do not answer questions about a person’s ability to safely perform the instrumental tasks required for independent daily life.
Chaytor and Schmitter-Edgecombe (2003) described the constraints limiting the ecological validity of neuropsychological assessments as including the disagreement in definitions of executive functioning; conducting tests in a static testing environment; measuring behaviors during a short period of time, therefore evaluating a limited sample of behaviors; not measuring compensatory strategies; and not acknowledging other factors, such as physical disabilities, emotions, and premorbid functioning. The EFPT offers an alternative approach to identifying executive dysfunction, one that can offer both the description of higher-order cognitive functions and a clinically useful tool that indicates the level of support a person needs to perform a task. Occupational therapists often are asked to provide an opinion on whether someone can live alone or be left alone during the day. The person’s performance on the EFPT offers an empirical underpinning to a decision that some level of assistance will be needed.

The moderate relationship between the FIM and the EFPT also suggests that similar constructs are being tapped, but the FIM cannot fully capture the complexity of performance. The FIM does not answer questions about the person’s ability to perform instrumental tasks or the level of help needed to perform them. The substantial relationship with the FAM does support the importance of using the EFPT to determine the performance needs of people with stroke as they perform IADL. Information collected will assist with family education.

Future Areas for Study

We will continue to add participants with more severity to more directly address the issues of both motor and cognitive impairment. With more participants, we will be able to do item analysis to determine the types of cognitive support needed by people with certain stroke syndromes to perform the activities that allow them to be safe in their instrumental activities at home.

Conclusion

Our study supports the use of the EFPT as a performance-based test of executive function that can be administered as a top-down assessment to build treatment plans in a home or clinic. It offers a client-centered approach that makes it possible for the clinician to consider safety and determine independence by observing the person performing a task. During the administration of the EFPT, it is possible to gain insight into environmental barriers that may limit performance and also identify strategies that can be used to support performance at home. It also provides the occupational therapist with information about the executive function issues that will interfere with daily life performance. The EFPT is a tool that occupational therapists can use to determine capacity and guide interventions with people with stroke and other chronic neurological conditions.

Acknowledgments

We recognize the following people for helping with the conceptualization of this assessment: Noomi Katz, PhD, OT; Adina Harman-Maier, PhD, OT; Deirdre Dawson, PhD, OT(C); Heidi Shamba, MD; and Jen Murawski, MSOT. This study was funded by James S. McDonnell Foundation Grant No. 21002032.

References


The American Journal of Occupational Therapy 453


