Factor Analysis of Nineteen Executive Function Tests in a Healthy Adult Population†

Renee Testa1,2,*, Pauleen Bennett1, Jennie Ponsford1

1Department of Psychology and Psychiatry, Monash University, Melbourne, Australia
2Melbourne Neuropsychiatry Centre, The University of Melbourne, Melbourne, Australia

*Corresponding author at: Melbourne Neuropsychiatry Centre, The University of Melbourne and Melbourne Health, Alan Gilbert Building, 161 Barry Street, Carlton South, 3053 Melbourne, Australia. Tel.: +61-3-8345-1303; fax: +61-3-8345-0599.
E-mail address: rtesta@unimelb.edu.au (R. Testa).

Accepted 10 December 2011

Abstract

Although there has been some progress in identifying the range of skills that comprise the executive neurocognitive system, their assessment has proved to be a challenge. Operationalization of executive functions (EFs) may be progressed by identifying the cognitive constructs that underlie EF test performance via principal components analysis. The underlying factor structure of 19 EF tests was examined in a non-clinical sample of 200 adults (mean age = 30.8 [18–64] years); the sample comprised 97 men. Findings revealed only weak correlations between various measures derived from the EF tests. Exploratory factor analysis revealed a model comprising six independent factors, consistent with previous reports, describing the functions of the EF system. The factors comprised: Prospective Working Memory, Set-Shifting and Interference Management, Task Analysis, Response Inhibition, Strategy Generation and Regulation, and Self-Monitoring and Set-Maintenance. Results confirm the diverse and heterogeneous nature of EFs and caution against conceptualizations that underestimate its complexity. Furthermore, variability within the “normal” executive system is evident, and further research is required to understand executive functioning in healthy populations.

Keywords: Executive function; Executive dysfunction; Frontal lobe; Neuropsychological tests; Cognition

Introduction

In the last decade, there has been rapid progress in regards to our understanding and conceptualization of the cognitive skills that comprise the executive system (Anderson, 1998; Chayer & Freedman, 2001). Neuropsychological investigations have significantly contributed by demonstrating the multidimensional and dissociable nature of executive functions (EFs) in both healthy non-clinical and clinical individuals (Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Duncan, Johnson, Swales, & Freer, 1997; Robbins et al., 1998). Despite this, debate has continued as to what is precisely meant by the term “executive function” and what criteria are to be used to assess its component skills. The skills assessed by tasks purported to measure executive functioning are also not well understood. As such, support for a model that is able to guide a comprehensive assessment has not been established (Graffman, 1995; Testa & Pantelis, 2009).

Three main approaches have proved informative for developing our current conceptualization of executive functioning. The first and most common approach involved lesion studies describing evidence from single case (de Oliveira-Souza, Moll, Moll, & de Oliveira, 2001; Eslinger, Flaherty-Craig, & Benton, 2004), and group studies of patients with frontal lobe (FL) damage (Baddeley, Della Sala, Papagno, & Spinnler, 1997; Godbout, Grenier, Braun, & Gagnon, 2005; Hornak et al., 2004; Jacobs & Anderson, 2002; Manes et al., 2002; McDonald, Delis, Norman, Tecoma, & Iragui-Madozi, 2005). This work provided insight...
into the functional specialization of EFs within frontal subregions, including the dorsolateral prefrontal cortex (DLPFC) and orbitofrontal cortex (OFC) (Baddeley & Della Sala, 1996; Bechara, Damasio, & Damasio, 2000; Duncan & Owen, 2000; Goldman-Rakic, 1987; Levine et al., 1998; Nauta, 1971; Passingham, 1985; Shallice, Burgess, & Frith, 1991; Stuss & Levine, 2002).

The DLPFC has been implicated in skills required for planning, decision-making and problem-solving, including strategy generation; goal formation and selection; monitoring and sequencing of behavior; anticipatory set; task analysis; decision-making; set-shifting; and abstract thinking (Aron, Robbins, & Poldrack, 2004; Burgess, 2000; Denckla, 1996; Fuster, 1991; Goldman-Rakic, 1987; Lezak, 1993; Norman & Shallice, 1986; Stuss & Benson, 1986; Tranel et al. 1994). In contrast, the OFC is thought to mediate social, affective, motivational, and personality aspects of an individual’s functioning, including the capacity for emotional and behavioral self-regulation, and integration of subjective experiences required for self-awareness and individuality (Cicerone & Tunenbaum, 1997; Stuss & Levine, 2002). More recently, Stuss (2007) has proposed a model of EF that describes four functional domains of EFs attributed to specialized frontal regions. These include executive cognitive functions; behavioral and emotional self-regulatory functions; activation-regulating functions; and metacognitive processes. In Stuss’s (2007) proposal, executive cognitive functions are largely mediated by the DLPFC and relate to skills of planning, monitoring, activating, switching, and inhibition. Behavioral self-regulatory skills involve ventral prefrontal and limbic regions, and are responsible for emotional processing. Activation-regulating functions include the ability to initiate and sustain appropriate behavioral responses to attain goals, and implicate the anterior cingulate and superior frontal regions. Metacognitive processes include skills related to personality, social cognition, consciousness and self-awareness, and are mediated by the frontal poles.

Secondly, evidence from neuroimaging studies (e.g. functional, structural, neurochemical) (Smith & Jonides, 1999; Szameitat, Schubert, Muller, & Von Cramon, 2002; Tamm, Menon, & Reiss, 2002; Taylor et al., 2004; Tekin & Cummings, 2002) has suggested that more detailed conceptualizations of the executive system are required, including evidence that executive dysfunction can occur following damage to non-frontal regions (Fassbender et al., 2004; Himanen et al., 2005; Roth & Saykin, 2004; van der Werf, Prins, Jongen, van der Meer, & Bleijenberg, 2000). Advances in brain imaging technologies have highlighted the inadequacy of a strict localizationist approach and the overly simplistic view of equating FLs and EFs (e.g. Blakemore & Choudhury, 2006; Charlton et al., 2006; Gothelf, Furfaro, Penniman, Glover, & Reiss, 2005; Holmes et al., 2005; Zimmerman et al., 2006). Both clinical and imaging studies have implicated EFs in non-frontal regions, various neural networks and many subordinate cognitive processes (Arnon et al., 2004; Benke, Delazer, Bartha, & Auer, 2003; Carter, Botvinick, & Cohen, 1999; Storey, Forrest, Shaw, Mitchell, & Gardner, 1999; Stuss & Alexander, 2000).

The third approach has incorporated the use of factor analytical techniques to identify constructs that underlie EF test performance in clinical and healthy populations. Despite this more psychometrically directed approach, investigations have reported diverse findings (Amieva, Phillips, & Della Sala, 2003; Bennett, Ong, & Ponsford, 2005; Burgess et al., 1998; Chan, 2001; Robbins et al., 1998). For example, Shute and Huertas (1990), Robbins and colleagues (1998), Pineda and Merchant (2003), and Bennett and colleagues (2005) demonstrated fractionation of the executive system in clinical and unimpaired samples. Robbins and colleagues (1998), using a large healthy sample, reported a four-factor solution using the Cambridge Neuropsychological Test Automated Battery (CANTAB) and the Tower of London (TOL) task. There were several cross-loadings, and derived factors were thought to represent Planning and Spatial Working Memory (SWM) (TOL and SWM), Attentional Set-Shifting (Interdimensional-Extradimensional or IDE task and the TOL), Strategic Aspects of Executive Functions (TOL and spatial working memory task), and Mnemonic Aspects of the Spatial Working Memory task. The solution accounted for 62.2% of the variance. Shute and Huertas (1990) investigated a smaller group of adults (n = 58) using measures purported to be sensitive to FL functioning (Category Test, Wisconsin Card Sort Task [WCST], Trail Making Test [TMT], Digit Symbol–WAIS [Wechsler Adult Intelligence Scale] and four experimental tasks), as well as a measure of operational reasoning (Piagetian Shadows Task). Their factor analysis produced four principal factors that accounted for 70% of the variance. A strong relationship between the operational reasoning task and the Category Test, TMT-B and the WCST was also identified, where these variables all loaded, although not exclusively, on the first principle factor in the solution. Pineda and Merchant (2003) identified five factors in a normal population using measures derived from four tests, but only listed the measures loading on each factor and did not offer an interpretative analysis of all findings. The factors were represented by the Stroop Errors measure, the Stroop Time measure, the TMT, Verbal Fluency and the WCST (the WCST factor was proposed to represent Organization and Flexibility). The solution also appeared to be stable and explained a high percentage of the shared variance (74.9%).

The identified factors and clinical interpretation has varied considerably for each factor analytic investigation, with results influenced by the limited number and type of executive measures included. A comprehensive portrayal of the executive system likely requires the inclusion of a larger number of measures assessing the full breadth of EFs. Large sample sizes to ensure the validity and stability of the solution in factor analysis are also required (Fabrigar et al., 1999). As it is often difficult to
administer a wide range of tests to large clinical populations, factor analytic investigations often fail to meet these statistical requirements, so that the generalizability of findings can be limited.

One challenge to the development of an EF model guiding assessment has been the wide range of measures of EF used in clinical practice. The majority of these have not been developed in direct relation to any of the EF theories or localization studies mentioned earlier, leaving clinicians with no clear guidelines as to what they are measuring when administering a given executive measure. The ability to administer well-designed and validated EF measures is necessary to progress on the operationalization of EFs. A structured evaluation of EF tests would help to clarify the construct validity of each measure, and assist in understanding the cognitive skills underlying each measurement. Also of relevance is the type of population. Predominantly, clinical samples are examined to infer a model of “healthy” executive functioning. However, relying on dysfunction as an indirect measure of unimpaired function may be inappropriate given that a linear relationship between a region’s function and dysfunction may not exist (Miller, 1990; Pantelis et al., 2003). In consideration, research characterizing EF variability within healthy, unimpaired populations may be necessary to develop a “gold standard” model of EF with which clinical findings can be compared.

In summary, the adoption of a range of investigative techniques has enabled a better understanding of EFs and the complex neurocognitive system that they comprise. Despite this large body of research, it is still difficult to identify which skills comprise EFs and how specific executive skills relate to or can be differentiated from one another in neuropsychological assessment. Whereas some previous factor analytic studies have been conducted, these have employed a limited number of tests and modest sample sizes, limiting the validity and generalizability of their findings.

The present study therefore aimed to examine the cognitive skills underlying a broad range of commonly used EF measures used in clinical and research settings; and explore the relationship between latent factors derived from these measures using an exploratory factor analytic approach in a large healthy population, in order to inform clinical assessment (Ardila, 2008; Denckla, 1994, 1996; Eslinger, 1996; Lezak, 1995).

Method

Participants

Two hundred adults were recruited from the general community via word of mouth and advertising within work organizations. Exclusion criteria included age <18 or > 65 years, and any past history of developmental or acquired brain injury. The participants comprised 97 men and 103 women, with a mean age of 30.8 years (range = 18–64 years, SD = 9.14). The percentage of participants in each age bracket was: 18–19 years (2.5%), 20–29 years (50%), 30–39 years (36%), 40–49 years (5.5%), 50–59 years (3.5%), and 60–65 years (2.5%). The mean estimated intelligence, as measured by the National Adult Reading Test (NART; Nelson, 1982) was 107.97 (range = 99.93–116.01), indicating that the sample was placed well within the average range. Fifty-four had completed or undertaken some secondary level education, 143 had completed or commenced tertiary courses, and three had completed or commenced TAFE (technical or trade) courses. Information obtained from the Australian Bureau of Statistics (ABS, 2001) indicated that the sample was representative of the general Australian population in terms of participants’ age and gender. Direct comparisons regarding education level were not possible, as detailed information (i.e. specific number of years and degrees achieved) was not recorded.

Materials

Nineteen neuropsychological tests were administered to all participants. The EF tests were drawn from an exhaustive literature review of 31 executive measures cited as being sensitive to executive deficits. All tests were assessed on four criteria, including: (i) how commonly the tests were used, (ii) if the tests were favorably reviewed in the literature, (iii) the psychometric properties of the measures, and (iv) the skills that the tests were purported to assess, in order to ensure that a wide range of executive skills was included. Each test was allocated a score ranging from −2 (fails to meet criteria), 0 (represents a neutral view or lack of data in the literature), or 2 (clearly meets the specific criteria), resulting in scores from −8 to 8. Tests assigned a score of 4 or above were included, with the aim of being more inclusive than exclusive. Two clinical neuropsychologists (R.T. and P.B.) independently evaluated each test according to the designated criteria. There was agreement between the raters as to which tests should be included within the study.

The selected measures included: Porteus Mazes (Porteus, 1965), Random Number Generation (RNG; Brugger, 1997), Six Elements Test (SET; Wilson et al., 1998), Key Search Task (Wilson et al., 1998), Zoo Map Test (Wilson et al., 1998), Stroop Test (Golden, 1978), Hayling Sentence Completion Test (Burgess & Shallice, 1996a), Brixton Spatial Anticipation Test (Burgess & Shallice, 1996b), Similarities (WAIS-III; Wechsler, 1997), Twenty Questions Test (Mosher & Hornsby, 1966),
were removed from the data and replaced with a score corresponding to that value (Tabachnick & Fidell, 1996). Although some
variables were found not to conform to normal distributions, they were not transformed due to the large sample size employed
in this study, and because EFA techniques are robust to violations of normality (Floyd & Widaman, 1995; Gorsuch, 1983;
variables.

Statistical Analysis

Previous EF studies have adopted both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Although
both models aim to identify the underlying structure of a set of variables, EFA imposes few restrictions on the relationship
between the variables and the number of factors identified (Fabrigar et al., 1999), whereas CFA is underpinned by theoretical
or empirical hypotheses. EFA was deemed most appropriate for the present study, as there was insufficient information avail-
able regarding the executive measures and their potential inter-relationships in a healthy population from which a defensible or empirical hypotheses. EFA was deemed most appropriate for the present study, as there was insufficient information avail-
able regarding the executive measures and their potential inter-relationships in a healthy population from which a defensible

All variables were examined for outliers and violations of normality. Outliers that were greater than 3.29 SD from the mean
were removed from the data and replaced with a score corresponding to that value (Tabachnick & Fidell, 1996). Although some
variables were found not to conform to normal distributions, they were not transformed due to the large sample size employed
in this study, and because EFA techniques are robust to violations of normality (Floyd & Widaman, 1995; Gorsuch, 1983;
Tabachnick & Fidell, 1996; de Vaus, 2002).

In order to examine the cognitive components underlying performance on the selected executive tasks, a principal compo-
nents analysis (PCA) with varimax rotation was used (Floyd & Widaman, 1995). The selection of dependent variables was
done with careful consideration (Fabrigar et al., 1999; Nunnally & Bernstein, 1994) in order to provide a solution that best
represented the EF constructs assessed by the measurements (Nunnally & Bernstein, 1994). Specific guidelines to ensure uni-
formity and inclusion of the most informative and worthwhile variables were developed, as outlined subsequently.
Guidelines for Variable Selection and Data Reduction

Given the large number of potential variables, a systematic approach for variable selection was devised (Table 1). All profile scores or scaled scores were eliminated as such “overlapping” variables essentially reflect an identical measure (Kline, 1994). They also have a limited range and are therefore poor discriminators, which can obscure real differences within a sample (Floyd & Widaman, 1995; Kline, 1998; de Vaus, 2002). In addition, three other statistical methods were adopted. The first technique examined correlations between each of the test variables. When variables were highly correlated (correlations > 0.7), all but one of the scores was eliminated; selecting the variable most commonly used in previous research (de Vaus, 2002). The second data reduction technique involved employing a PCA; this occurred when a particular test (WCST and CNT) provided values that represented different facets of EF, and deriving one overall score was not possible or would have resulted in the loss of important information (Tabachnick & Fidell, 1996). The third data reduction technique included examination of the correlations between the variables and other tests proposed to assess similar executive skills, given that it is easier to identify patterns of correlations when the degree of inter-correlation between a set of variables is maximized (Nunnally & Bernstein, 1994). The variable found to have the highest correlation with the other executive measures was selected for inclusion. This technique was applied if the (two) previously described methods were not able to sufficiently reduce the data.

Results

Data Reduction of the WCST and CNT

Results from the PCA analysis of WCST scores indicated that a two-factor solution accounted for 78.19% of the total variance. Factor scores from this analysis were subsequently used (i.e. variable names WCSTFA and WCSTFB). Regarding interpretation of factor scores, the majority of measures from the task, including the number of correct responses, the number of categories achieved, and the number of conceptual level responses, were included within the first factor. This factor likely

Table 1. Variable label for each executive function measure selected for the statistical analyses.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Measure used</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porteus Maze</td>
<td>Total Number of Type One and Two Errors (PMTOTERR)</td>
<td>4.00</td>
<td>0.40</td>
<td>0.81</td>
</tr>
<tr>
<td>Hayling Test</td>
<td>Part B minus Part A Total Time (HAYB-AT)</td>
<td>119.00</td>
<td>15.15</td>
<td>2.61</td>
</tr>
<tr>
<td>Hayling Test</td>
<td>Part B: Total Category A Errors plus Total Category B Errors (HAYBAERR)</td>
<td>12.00</td>
<td>2.11</td>
<td>2.61</td>
</tr>
<tr>
<td>20 Questions Test</td>
<td>Percentage of Constraint Seeking Questions (TWPERCS)</td>
<td>77.37</td>
<td>66.54</td>
<td>15.19</td>
</tr>
<tr>
<td>20 Questions Test</td>
<td>Percentage of Pseudo-Constraint Seeking Questions (TWPERPCS)</td>
<td>37.98</td>
<td>6.21</td>
<td>9.00</td>
</tr>
<tr>
<td>Wechsler Adult Intelligence Scale–Similarities</td>
<td>Raw Score (SIM)</td>
<td>23.00</td>
<td>22.45</td>
<td>4.55</td>
</tr>
<tr>
<td>Stroop Test</td>
<td>Color Word Score (STPCWS)</td>
<td>53.00</td>
<td>46.33</td>
<td>9.92</td>
</tr>
<tr>
<td>Brixton Test</td>
<td>Total Number of Errors (BRIXERR)</td>
<td>24.00</td>
<td>12.40</td>
<td>4.48</td>
</tr>
<tr>
<td>Cognitive Estimates Test</td>
<td>Total Z Score Average from the Mean Score (CETZAVMC)</td>
<td>2.42</td>
<td>-0.0019</td>
<td>0.35</td>
</tr>
<tr>
<td>Key Search Test</td>
<td>Total Raw Score (KS)</td>
<td>14.00</td>
<td>12.48</td>
<td>3.53</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>Time for Part B minus Time for Part A (TMTBADIF)</td>
<td>97.00</td>
<td>32.88</td>
<td>18.62</td>
</tr>
<tr>
<td>Six Elements Test</td>
<td>Total Number of Subtasks where Rules were Broken (SETRUTOT)</td>
<td>5.00</td>
<td>0.31</td>
<td>0.90</td>
</tr>
<tr>
<td>Six Elements Test</td>
<td>Total Number of Switches (SETSWITC)</td>
<td>46.00</td>
<td>8.28</td>
<td>7.12</td>
</tr>
<tr>
<td>Zoo Map Test</td>
<td>Zoo Map One Raw Score (ZMORS)</td>
<td>6.00</td>
<td>0.64</td>
<td>1.22</td>
</tr>
<tr>
<td>Verbal Fluency</td>
<td>Total Number of Words (VFTOTWDS)</td>
<td>68.00</td>
<td>44.14</td>
<td>11.24</td>
</tr>
<tr>
<td>Verbal Fluency</td>
<td>Total Number of Errors (VFTOTERR)</td>
<td>7.00</td>
<td>1.45</td>
<td>1.60</td>
</tr>
<tr>
<td>Contingency Naming Test</td>
<td>Factor A Score (CNTFA)</td>
<td>8.85</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Contingency Naming Test</td>
<td>Factor B Score (CNTFB)</td>
<td>5.37</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Animal Fluency</td>
<td>Number of Words in Size List (AFWRDS1)</td>
<td>17.00</td>
<td>13.04</td>
<td>3.04</td>
</tr>
<tr>
<td>Animal Fluency</td>
<td>Number of Size Errors (AFWERRS)</td>
<td>5.00</td>
<td>1.07</td>
<td>1.20</td>
</tr>
<tr>
<td>Concept Generation Test</td>
<td>Number of Categories Achieved (CGTNO.CT)</td>
<td>5.00</td>
<td>4.23</td>
<td>1.27</td>
</tr>
<tr>
<td>TOL–Revised</td>
<td>Total Time Taken for all Trials (TOLTTC)</td>
<td>1254.00</td>
<td>394.58</td>
<td>252.12</td>
</tr>
<tr>
<td>TOL–Revised</td>
<td>Total Trials Correct (TOLTC)</td>
<td>11.00</td>
<td>11.31</td>
<td>2.56</td>
</tr>
<tr>
<td>WCST</td>
<td>Factor A Score (WCSTFA)</td>
<td>4.55</td>
<td>0.14</td>
<td>0.88</td>
</tr>
<tr>
<td>WCST</td>
<td>Factor B Score (WCSTFB)</td>
<td>6.20</td>
<td>0.12</td>
<td>0.85</td>
</tr>
<tr>
<td>Random Number Generation</td>
<td>Repetition (RNGR)</td>
<td>35.00</td>
<td>17.94</td>
<td>3.07</td>
</tr>
<tr>
<td>Random Number Generation</td>
<td>Seriation (RNGS)</td>
<td>13.63</td>
<td>9.00</td>
<td>2.62</td>
</tr>
<tr>
<td>Random Number Generation</td>
<td>Cycling (RNGC)</td>
<td>15.92</td>
<td>6.47</td>
<td>3.14</td>
</tr>
<tr>
<td>e-Dysexecutive Questionnaire</td>
<td>Total Score (e-DEX)</td>
<td>143.00</td>
<td>63.92</td>
<td>27.54</td>
</tr>
</tbody>
</table>
represents the reasoning, problem-solving element of the task. In contrast, the second factor included the trials to complete first category, failure to maintain set and total correct, and reflected abilities required for set maintenance or “staying on task.”

The CNT produced a five-factor solution that accounted for 64.64% of the variance. Examination of the scree plot indicated, however, that a three-factor solution was likely a better fit. This accounted for 46.30% of the variance. Factor-derived scores from this analysis were subsequently used in the final factor analysis (i.e. variable names CNTFA, CNTFB and CNTFC). In terms of factor score interpretation, the first factor comprised scores from the baseline Trials 1 and 2. The second factor included the number of self-corrections made by participants on Trials 3 and 4. The last factor comprised the number of correct responses and time for completion on Trial 4, as well as the number of correct responses on Trial 1.

Analysis of Correlations between the EF Measures

Pearson’s product–moment correlations were conducted using the test variables selected. For ease of interpretation, test variables were grouped according to each of the commonly cited executive domains (as reported earlier) that they purportedly assess. Bonferroni corrections were also applied (p-value < .0001) because of the large number of correlations performed. In the original (uncorrected) matrix, there were 55 intra-task and four inter-task correlations that were significant at the < .01 level. As predicted, application of the Bonferroni correction resulted in an overall reduction in the number of significant correlations. A considerable number of correlations remained significant however, but were generally weak. Within the matrix, 31 inter-task correlations were significant at the $p < .0001$ level, as well as 11 intra-task correlations. In consideration of the multiple comparisons conducted within the dataset, there is an increased likelihood of experiment-wise Type 1 error being present. However, given the strict $p$ value set when Bonferroni adjustments were conducted, it was not considered possible to further reduce the set significance level and further adjust for multiple comparisons. Further, an omnibus null test (Cohen, 2000) was conducted on the correlations that did not share method variance. It was found that, despite the large number of correlations conducted, the chi-square value was significant at $p < .001$, where $\chi^2 = 1264.74$, $df = 406$, and the critical level of significance = 553.455.

Although the number of correlations conducted is too extensive to report in this paper, in summary, the majority of correlations that were identified occurred between variables that were derived from the same task. Correlations between different executive tasks were weak to moderate and limited in number, demonstrating the independence of the tasks. If these tests are valid and reliable measures of the executive system, and the system comprises dissociable components, then higher correlations between measures that are said to assess similar executive constructs would be expected.

PCA of EF Measures

Guidelines for selecting the best solution were obtained from Fabrigar et al. (1999). All selected test variables were subject to a PCA with varimax rotation (Fabrigar et al. 1999). The solution accounted for 62.99% of the variance and produced 11 factors, using Kaiser’s criterion of eigenvalues greater than one. Evidence for the dissociability of the solution was supported by the unrotated solution, which indicated that the first unrotated principal component was comparable in size with other components. However, the scree plot indicated that either five or six factor solutions may be more appropriate; the solution was therefore re-run restricting the solution. In order to obtain the optimum solution, measures that loaded below 0.4 were removed and the analysis was conducted again following the extraction of the specified number of variables (Floyd & Widaman, 1995; Tabachnick & Fidell, 1996).

Examination of the results from both analyses demonstrated that the six-factor solution produced the best result, in that it was well-defined and offered a simple structure with only one measure loading across the factors. It also accounted for a greater percentage of the variance, being 46.54% compared with the 42.23% for the five-factor solution. Table 2 shows the six-factor solution with loadings less than 0.4 excluded. The percentage of variance accounted for by each of the individual factors (with eigenvalues presented in parentheses) was 8.43% (2.10), 8.19% (2.05), 8.09% (2.02), 7.57% (1.88), 7.28% (1.82) and 7.02% (1.76), respectively.

Factor One comprised four measures. Two were from the same test (TOL-R) and the others were the Porteus Maze and the RNG Seriation measure. These variables loaded strongly on the factor, with loadings ranging from 0.50 to 0.81. Executive Factor One was thought to represent Prospective Working Memory. Executive Factor Two comprised three measures: the Stroop, TMT and the CNT (F1) factor score. All three variables demonstrated comparable and strong loadings on this factor, thought to represent Set-Shifting and Interference Management. Executive Factor Three comprised five measures. Three of the measures loaded moderately on this factor: the Zoo Map test, the Brixton test, and the WCST (FA) measure. This factor was thus thought to assess Task Analysis. The fourth and fifth measures, which were found to load relatively more weakly, were from the CET and the CNT (F2). Executive Factor Four involved three measures. Two loaded strongly
on this factor and were derived from the Hayling test, while the third measure, which loaded more moderately, was the Twenty Questions Test Percentage of PCS questions. This factor was thought to reflect Response Inhibition. The CNT (FB) factor score was also found to load on this factor, although it contributed slightly more to the third factor in the solution. Executive Factor Five comprised four measures, three of which were found to load strongly. Two of the measures were derived from the SET and one measure from the Verbal Fluency task. In comparison, a weaker loading was observed for the last measure, the Animal Fluency (number accurately recalled in size order) measure. This factor represented Strategy Generation and Regulation.

Five comprised four measures, three of which were found to load strongly. Two of the measures were derived from the SET and one measure from the Verbal Fluency task. In comparison, a weaker loading was observed for the last measure, the Animal Fluency (number accurately recalled in size order) measure. This factor represented Strategy Generation and Regulation.

Executive Factor Six also comprised four measures. These were the WCST (F2) and CNT (F3) scores, in addition to two Fluency (number accurately recalled) measures. This factor represented Strategy Generation and Regulation.

One measure from the Verbal Fluency task. In comparison, a weaker loading was observed for the last measure, the Animal Fluency (number accurately recalled in size order) measure. This factor represented Strategy Generation and Regulation.

Executive Factor Six also comprised four measures. These were the WCST (F2) and CNT (F3) scores, in addition to two Fluency (number accurately recalled) measures. This factor represented Strategy Generation and Regulation.

To further validate the findings, the sample population was divided into two groups and the factor analysis was again conducted with a sample of 100 participants in each group; results comparable with the original solution were identified, adding further support to the stability of the solution reported.

### Discussion

The aim of the present investigation was to characterize the relationship between current neuropsychological tests and executive cognitive skills. This is informative for developing a conceptualization of EFs, as measured by the assessment tasks used to assess this cognitive system. Delineating the inter-relationships between executive measures can also assist in the interpretation of neuropsychological findings.

Examination of the correlations between the test variables generally demonstrated very weak relationships. The independence of the executive skills assessed by each of the measures, and the multifactorial nature of the measures likely contributes to this finding. Successful EF test completion is believed to require a number of executive and non-executive skills (Burgess,

---

Table 2. Rotated component matrix loadings of orthogonally rotated factors extracted by principal components factor analysis

<table>
<thead>
<tr>
<th>Neuropsychological measures</th>
<th>Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor 1</td>
</tr>
<tr>
<td>Tower of London (Total Time)</td>
<td>0.81</td>
</tr>
<tr>
<td>Random Number Generation (Seriation)</td>
<td>-0.66</td>
</tr>
<tr>
<td>Tower of London (Total Correct)</td>
<td>0.65</td>
</tr>
<tr>
<td>Porteus Mazes</td>
<td>-0.50</td>
</tr>
<tr>
<td>Contingency Naming Test (FA)</td>
<td>-0.03</td>
</tr>
<tr>
<td>Stroop Color Word Test</td>
<td>0.10</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>-0.01</td>
</tr>
<tr>
<td>Zoo Map Test</td>
<td>0.06</td>
</tr>
<tr>
<td>Wisconsin Card Sort Test (FA)</td>
<td>0.09</td>
</tr>
<tr>
<td>Brixton Test</td>
<td>-0.25</td>
</tr>
<tr>
<td>Contingency Naming Test (FB)</td>
<td>0.03</td>
</tr>
<tr>
<td>Cognitive Estimates Test</td>
<td>-0.09</td>
</tr>
<tr>
<td>Hayling Test (Total Time)</td>
<td>-0.01</td>
</tr>
<tr>
<td>Hayling Test (Error Score)</td>
<td>-0.12</td>
</tr>
<tr>
<td>Twenty Questions Test (% of PCS Qns)</td>
<td>-0.01</td>
</tr>
<tr>
<td>Six Elements Test (Switches)</td>
<td>0.01</td>
</tr>
<tr>
<td>Six Elements Test (Total Score)</td>
<td>-0.08</td>
</tr>
<tr>
<td>Verbal Fluency Test (Number of Words)</td>
<td>0.16</td>
</tr>
<tr>
<td>Animal Fluency Test (Number of Words)</td>
<td>-0.06</td>
</tr>
<tr>
<td>Random Number Generation (Repetition)</td>
<td>-0.02</td>
</tr>
<tr>
<td>Wisconsin Card Sort Test (FB)</td>
<td>0.13</td>
</tr>
<tr>
<td>Contingency Naming Test (FC)</td>
<td>0.09</td>
</tr>
<tr>
<td>Random Number Generation (Cycling)</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

Notes: FA = Factor Score A; FB = Factor Score B; FC = Factor Score C. Bold number indicate factor loadings ≥ 0.40.
approach is modified if it is not efficacious or in accordance with task demands. Underpinning this is the capacity to be cognitively flexible and self-regulate responses. This would ensure that the strategic approach is modified if it is not efficacious or in accordance with task demands.

For this reason, the adoption of a factor analytic approach was considered most informative. Factor analysis of the executive measures in this study demonstrated the existence of six dissociable components. Low correlations between the identified factors demonstrated their independence, supporting previous notions of a fractionated executive system (Duncan et al., 1997; Miyake et al., 2000). This study is one of the first investigations to establish a relationship between executive function constructs and a large group of specific neuropsychological measures.

Although it is acknowledged that interpretation of factors is difficult, the authors have discussed subsequently the common executive skills that may represent each of the identified constructs. This interpretation was based on the previous literature examining each of the EF tasks, which was initially reviewed in the test selection process. Additional investigation would be required to validate the factors identified.

Executive Factor One was considered representative of Prospective Working Memory. The TOL-R, Porteus Mazes and RNG are all tasks that require the ability to hold on line possible responses, and think or look ahead in the short-term using working memory capacity, so that the optimum behavioral response or solution can be selected.

Executive Factor Two was thought to represent Set-Shifting and Interference Management. Tasks in this construct require a good capacity to process, manage and shift between more than one element of a stimulus concurrently. The TMT requires participants to mentally shift between two sets of information (letters and numbers), and stay on task while coping with more than one task demand simultaneously (dual task). This is also a component skill needed in the CNT. Further, the Stroop Task requires the ability to efficiently shift attention from the most salient aspect of the stimuli (i.e. reading the word) to the more important aspect for successful task performance (i.e. the color in which the word is printed).

Executive Factor Three included measures from the Zoo Map test, the Brixton test, and the first derived factor score from the WCST (FA), in addition to more moderate contributions from the Cognitive Estimates Task (CET) and the CNT (FB). Overall measures included in this factor appear to represent Task Analysis. Successful performance requires good problem-solving skills including identifying and analyzing the task’s requirements. For example, the WCST, Zoo Map Test and Brixton Test require participants to understand the cognitive reasoning that underpins these tasks, and identify and implement the most appropriate behavioral approach. This includes a component of conceptual or abstract reasoning, in addition to analysis of the task requirements and rules. Successful performance of the CET is also dependent on these skills, requiring evaluation, judgment and reasoning.

Executive Factor Four was thought to represent Response Inhibition. The Hayling Test is a measure of response inhibition, including the ability to stop or inhibit pre-potent responding. Similarly the Twenty Questions Test (PSC) variable, included in this construct, is considered a measure of inhibiting inappropriate responding. The CNT (FB) factor score additionally loaded on this factor, although it contributed more strongly to Executive Factor Three. This cross-loading is not unexpected given that CNT (FB) represents number of self-corrections exhibited, which is related to the participant’s capacity to inhibit incorrect responses. It may have been expected that the Stroop task would have loaded within this factor given that it is said to assess verbal response inhibition. However, given that it loaded strongly and exclusively within Factor 2, it appears to be primarily assessing an individual’s capacity to deal with more than one aspect of task stimuli concurrently rather than taxing their ability to inhibit a response. Further, it also differs in that the self-generation of a response is not required (comparable with tasks in Factor Four).

Executive Factor Five encompasses Strategy Generation and Regulation. This is a multidimensional construct, which is proposed to comprise the ability to select and implement strategies to produce the optimum task response. The adoption of a strategy facilitates successful performance for tasks within this construct. For example, on the Animal Fluency task, use of a strategy such as listing all African animals will ensure that the largest number of items can be generated. Underpinning this is the capacity to be cognitively flexible and self-regulate responses. This would ensure that the strategic approach is modified if it is not efficacious or in accordance with task demands.

Executive Factor Six comprised Self-Monitoring and Set-Maintenance (Factor 6). The WCST, RNG and CNT variables included within this factor may represent the ability to identify incorrect responding or inappropriate behavior, and modify behavior in response to feedback. This can be achieved by self-monitoring performance, as required in the RNG to ensure random production of responses, or the CNT to ensure that the participant is adhering to task rules. It can also be provided by environmental input, such as examiner response feedback in the WCST.

It is acknowledged that each factor represents an integrated and multi-dimensional set of skills. Dysfunction of the executive system could therefore be attributable to any of the identified skills of which it is comprised. Further dissociation of each construct is likely possible, and although beyond the scope of this paper, will be important for better understanding the skills that
contribute to these constructs and the tests used to assess EF. It is pertinent that both clinicians and researchers acknowledge the multifactorial nature of these tasks when they are administered.

Previous factor analytic studies have not been sufficiently comprehensive to provide an overview of EF tests and the skills they assess. The strength of the current investigation lies in the uniquely large number of measures included. In addition, the current findings are supported by the degree of congruence between the literature reports of EF skills earlier discussed and the constructs identified. In previous lesion studies, executive skills including strategy generation; inhibition, goal formation, and selection; monitoring of behavior; and problem-solving are all directly comparable with the constructs in this investigation (Aron et al., 2004; Burgess, 2000; Denckla, 1996; Fuster, 1991; Goldman-Rakic, 1987; Lezak, 1993; Norman & Shallice, 1986; Stuss & Benson, 1986; Tranel et al. 1994). Notably, these skills are largely attributable to the DLPC rather than the OFC, which lacks standardized assessment in neuropsychological practice, and were therefore not identified in this investigation.

Findings of the current study also support the EF model of Stuss (2007) that describes four components of executive functioning. The identified factors are directly comparable with two of Stuss’ theoretical constructs. Executive cognitive skills (planning, monitoring, activating, switching, and inhibition) can be directly related to this study’s constructs of Set-Shifting and Interference Management, Task Analysis, Self-Monitoring and Set-Maintenance, and Response Inhibition; whereas, activation-regulating functions (ability to initiate and sustain appropriate behavioral responses to attain goals) are comparable with the Strategy Generation and Regulation and Prospective Working Memory factors in this study. Although these executive factors are thought to be mediated by the DLPC, the remaining two components of Stuss’ model involved in social, emotional processing, personality, and self-awareness (i.e. behavioral self-regulatory skills and metacognitive processes) implicate OFC regions. It is therefore not unexpected that these factors would not be comparable with the current findings.

Comparisons with previous exploratory factor analytic studies in healthy groups are also beneficial. Robbins et al.’s (1998) four-factor solution demonstrated similarities with the current findings: direct links between Robbins et al.’s Attentional Set-Shifting and Strategic Aspects of Executive Functions constructs, and this study’s findings of Set-Shifting and Interference Management and Strategy Generation and Regulation constructs were identified. Robbins et al.’s Planning and Spatial Working Memory factor can be partially related to the Task Analysis construct identified in this study (with the main component being the problem-solving/planning of tasks). Significant differences are evident between Robbins et al.’s Mnemonic Aspects of the Spatial Working Memory task, and the current study’s Prospective Working Memory, Response Inhibition, and Self-Monitoring and Set-Maintenance.

Direct comparisons are limited with the other investigations. Shute and Huertas’ (1990) analysis produced four factors comprising an operational reasoning task and the Category Test, TMT-B and the WCST which all loaded, although not exclusively, on the first principle factor in the solution. The TMT and WCST loaded on separate factors in the current study. Pineda and Merchan (2003) identified five factors derived from four tests, but listed only the measures loading on each factor and did not offer an interpretative analysis of all findings. The factors were represented by the Stroop Errors measure, the Stroop Time measure, the TMT, Verbal Fluency, and the WCST (the WCST factor was proposed to represent Organization and Flexibility). In the current study the Stroop task and TMT were placed within the same factor, but the WCST and Verbal Fluency loaded on different constructs. The construct on which the WCST loaded was defined as Task Analysis, and indicative of reasoning and problem-solving skills.

Given the significant methodological differences between the earlier mentioned and current investigations, it is evident that a comprehensive description of skills underlying EF performance is unlikely to emerge from investigations using a relatively small number of tests. Comparisons between this and future investigations adopting a similarly large number of tasks will be useful to determine whether executive tests are sensitive to comparable executive components in other clinical and healthy groups.

The results of this study have significant implications on the clinical assessment of executive functioning. They have confirmed that the EF system is multifactorial, but also that an expanded number of independent executive skills may exist. In order to obtain a comprehensive assessment of this system, a larger and more comprehensive battery may need to be used to provide a comprehensive picture of executive difficulties. The failure of tests to identify executive dysfunction may not only be a result of the poor psychometrics of the test, but also reflect the possibility that the administered test is not sensitive to the specific executive difficulty that the patient is experiencing.

The results are also informative for the clinician as they provide a greater understanding of how commonly used EF measurements are inter-related; that is the extent to which individual tests assess unique or overlapping aspects of EFs. Clinical neuropsychologists often rely on profile analysis of various EF test scores to understand the cognitive strengths and weaknesses of their clients; this is often based on their clinical experience, and knowledge of their tests, cognition and neuroanatomy. Any literature that can assist clinicians to understand what types of cognitive processes their clinical measures are actually assessing is useful to determine whether executive tests are sensitive to comparable executive components in other clinical and healthy groups.

The results of this study have significant implications on the clinical assessment of executive functioning. They have confirmed that the EF system is multifactorial, but also that an expanded number of independent executive skills may exist. In order to obtain a comprehensive assessment of this system, a larger and more comprehensive battery may need to be used to provide a comprehensive picture of executive difficulties. The failure of tests to identify executive dysfunction may not only be a result of the poor psychometrics of the test, but also reflect the possibility that the administered test is not sensitive to the specific executive difficulty that the patient is experiencing.

The results are also informative for the clinician as they provide a greater understanding of how commonly used EF measurements are inter-related; that is the extent to which individual tests assess unique or overlapping aspects of EFs. Clinical neuropsychologists often rely on profile analysis of various EF test scores to understand the cognitive strengths and weaknesses of their clients; this is often based on their clinical experience, and knowledge of their tests, cognition and neuroanatomy. Any literature that can assist clinicians to understand what types of cognitive processes their clinical measures are actually assessing is useful to determine whether executive tests are sensitive to comparable executive components in other clinical and healthy groups.
is regarded to be very important, and thought to be addressed to a degree by the current investigation. Further, understanding how non-clinical individuals perform on these tasks can assist the interpretation of the performance of clinical populations.

With regards to theoretical models of the EF system, the current findings highlight that the inconsistency of research findings may reflect methodological differences in the choice of tests used, often combined with small sample sizes. Studies utilizing only a few tests are unlikely to capture the full spectrum of executive skills, resulting in an oversimplification of the concept of EFs.

It is acknowledged that there are several limitations to this study. Recruitment did not specifically control for age, education, intelligence or employment history; and while using a normal population was considered a strength of this study, it would be informative to compare and validate the factor structure of EF measures in a neurologically impaired sample. If analogous executive factors are identified, research could be directed at investigating each of the independent constructs, and develop “purer” tasks that specifically measure the skill(s) comprising each. Clearly, this would facilitate test interpretation for both the clinician and the researcher. Given that previous research has demonstrated that the stability of factor solutions is dependent on the characteristics of the population being investigated (Chan, 2001; Pineda & Merchan, 2003; Robbins et al., 1998), understanding the profile of each EF factor structure within different clinical samples may be necessary.

In summary, this study offered the opportunity to examine the performance of a large healthy adult sample on an extensive range of EF tests. Results indicated the existence of six constructs including Prospective Working Memory, Set-Shifting and Interference Management, Task Analysis, Response Inhibition, Strategy Generation and Regulation, and Self-Monitoring and Set-Maintenance and identified the contribution of each test to the measurement of these constructs. This finding is consistent with current conceptualizations of EF in the literature, and confirms the notion that EF tests assess a diverse range of skills. The current study has thereby progressed our understanding of the relationship between specific EFs and their relationship with neuropsychological measures used in clinical and research practice. This study provides an important benchmark against which other investigations, attempting to obtain a more comprehensive portrayal of the executive system in healthy and clinical populations can be compared.

Conflict of interest: none declared.

References


