Improving the ecological validity of executive functioning assessment

Naomi Chaytor a, b, *, Maureen Schmitter-Edgecombe a, Robert Burt c

a Washington State University, Department of Psychology, Pullman, WA, United States
b Baylor College of Medicine, Department of Neurosurgery, Houston, TX, United States
c Sacred Heart Medical Center, Neuropsychology Program, Spokane, WA, United States

Accepted 13 December 2005

Abstract

The current study investigated ways to improve the ecological validity of the neuropsychological assessment of executive functioning through the formal assessment of compensatory strategies and environmental cognitive demands. Results indicated that the group of executive functioning tests (i.e., Trail Making Test, Wisconsin Card Sorting Test, Stroop, and Controlled Oral Word Association Test) accounted for 18–20% of the variance in everyday executive ability as measured by the Dysexecutive Questionnaire and Brock Adaptive Functioning Questionnaire. The addition of extra-test variables significantly increased the variance in everyday executive ability accounted for. The current study adds to the literature on the ecological validity of executive functioning assessment by highlighting the importance of extra-test variables when trying to understand the complex relationship between cognitive testing and real world performance.

Keywords: Executive functioning; Ecological validity; Neuropsychological assessment

Neuropsychologists are often perplexed by clients whose performance on tests of executive functioning is inconsistent with their everyday executive functioning abilities (Wilson, 1993). It is assumed that the impaired brain processes, which lead to poor performance on a neuropsychological test, will also lead to poor performance in other situations outside the test situation. In other words, it is assumed that neuropsychological tests have ecological validity. This is a key assumption, as poor performance that is only observed in the context of neuropsychological tests would be of limited clinical relevance. Surprisingly, there has been very little research investigating the accuracy of this assumption.

The current study sought to replicate previous research examining the ecological validity of executive tests, and to build upon this research by investigating ways to improve the ecological validity of executive functioning assessment.

Of critical importance to research on ecological validity is how everyday cognitive ability is measured. Previous ecological validity research has employed self and informant questionnaires, clinician rating scales, or observation of simulated everyday tasks. There is no perfect way to quantify true everyday cognitive ability, as any form of assessment is associated with a certain degree of error. For example, the observation of a client in a simulated setting could result...
in greater agreement with neuropsychological tests because the conditions of observation (artificial situation, stress of being evaluated) may be similar to the testing environment and may not accurately reflect the client’s typical everyday abilities. Likewise, relying on the brain-damaged patient to report on his or her own cognitive ability also poses problems with measurement error due to lack of insight or impaired memory. Thus, the majority of the studies of ecological validity of executive functioning assessment have relied on informant-based questionnaires and clinician rating scales, both because of relative ease of data collection and fewer systematic biases.

The literature investigating the ecological validity of neuropsychological tests of executive functioning has been inconsistent: with some studies demonstrating relatively robust relationships between test scores and everyday ability, and others failing to find significant relationships. In a large sample of neurologically impaired patients, Burgess, Alderman, Evans, Emmsie, and Wilson (1998) found significant individual correlations between many commonly used neuropsychological tests of executive functioning (e.g., Modified Wisconsin Card Sorting Task, Trail Making Test, Controlled Oral Word Association Test) and informant-based ratings of everyday executive skills (i.e., Dysexecutive Questionnaire; DEX), but not self-ratings. Several other studies investigating the ecological validity of executive functioning tests have also found significant relationships between neuropsychological test performance and clinician ratings on the Neurobehavioral Rating Scale (NRS) (Dimitrov, Grafman, & Hollnagel, 1996), informant ratings on the DEX (Evans, Chua, McKenna, & Wilson, 1997), clinician ratings of interactive skills and planning (Poole, Ober, Shenaut, & Vinogradov, 1999), parent ratings of hyperactivity and distractibility (Clark, Prior, & Kinsella, 2000), and teacher ratings and clinician observation of classroom behavior (Solanto et al., 2001). In contrast, other studies that have used similar neuropsychological tests (e.g., Wisconsin Card Sorting Task, Trail Making Test) and outcome measures (i.e., self and other-report DEX, clinician rating scale, self-report questionnaires) have found very little relationship between tests of executive functioning and measures of everyday cognitive skills (Amieva, Phillips, & Della Sala, 2003; Bogod, Matee, & MacDonald, 2003; Chan, 2001; Norris & Tate, 2000; Ready, Sierman, & Paulsen, 2001). Currently, however, it remains difficult to compare findings across studies as the studies have varied in the populations evaluated (e.g., Schizophrenia, Attention Deficit Disorder, Older Adults, General Neurological), the type and number of neuropsychological tests of executive functioning employed, and the outcome method utilized (e.g., clinician ratings, self-report, significant other report).

Despite study differences, even in those studies that have reported favorable ecological validity, the magnitude of the relationships tended to be moderate, ranging from .27 to .65, when significant. This means that a large amount of the variance in everyday executive skills remains unaccounted for. As pointed out by several authors (Chaytor & Schmitter-Edgecombe, 2003; Franzén & Arnett, 1997; Long & Collins, 1997; Long & Kibby, 1995; Sbordone, 1996; Sbordone & Guilmette, 1999; Silver, 2000), many characteristics of traditional neuropsychological assessment can pose problems for ecological validity research, effectively placing a limit on the amount of variance in everyday skills that can be accounted for by neuropsychological testing. These characteristics include the nature of the testing environment, incomplete agreement on what constructs the tests measure, the small sample of behavior observed during testing, and the inability to use compensatory strategies. There are also several noncognitive factors that can influence the relationship between test performance and everyday performance. These noncognitive factors include emotional problems, level of premorbid functioning, motor functioning, health problems, and varying environmental cognitive demands (Long & Kibby; Sbordone, 1997). Accounting for these variables, in addition to performance on executive functioning tests, may allow neuropsychologists to better predict everyday executive functioning. Although all of these factors have the potential to affect ecological validity, in this study we explore the contribution of environmental cognitive demands and compensatory strategies. These variables were chosen because of the potential for clinical application.

Environmental cognitive demands vary widely across individuals and may mediate the relationship between performance on neuropsychological tests and everyday functioning. Someone with an executive deficit may have no real world problems if his or her environment places little demand on this skill. Conversely, even minor executive deficits coupled with a highly demanding environment could cause extreme functional impairment. This highlights the need to assess the cognitive demands that would be required of any person to function in the person’s everyday environment and match these demands to cognitive test performance in order to accurately predict functional consequences (McCue & Pramuka, 1998; Sbordone, 1997; Sbordone & Guilmette, 1999). The cognitive deficit and the environment interact to produce behavior, and therefore both need to be assessed if ecological validity is to be demonstrated (Chelune, 1985; Goldstein, 1996; Heinrichs, 1990). To date, very little attention has been paid to the development of structured, objective methods for assessing environmental cognitive demands. Currently, subjective approaches are used, such as
clinician ratings of the environment as having high or low demand in a given cognitive domain (e.g., high memory demands, low executive demands; McCue & Pramuka). The inclusion of environmental demand assessment in studies of ecological validity has the potential to clarify the relationship between neuropsychological tests and everyday cognitive performance and possibly account for additional variance.

Another factor that has the potential to obscure the relationship between executive functioning test performance and everyday skills is use of compensatory strategies (Long & Kibby, 1995). An individual may use compensatory skills in everyday life, but is prevented from using them in the testing situation. Alternatively, the individual may fail to use cognitive strategies in everyday situations where it is expected or needed. Assessment of the individual’s everyday compensatory strategy use might add important information when predicting everyday executive skills.

The current study attempts to replicate previous work investigating the relationship between several commonly used neuropsychological tests of executive functioning and significant other reports of everyday executive function. Further, the current study will investigate ways to improve the ecological validity of the neuropsychological assessment of executive functioning. More specifically, to determine whether the formal assessment of compensatory strategies and environmental cognitive demands can account for additional variance in informant ratings of everyday executive skills beyond that accounted for by neuropsychological tests of executive functioning. The following are the specific research questions that were addressed in the current study:

1. Are executive functioning tests related to informant report of everyday executive functioning skills?
2. Does assessment of environmental cognitive demands account for additional variance in everyday executive functioning beyond that of the executive functioning tests alone?
3. Does assessment of compensatory strategies account for additional variance in everyday executive functioning beyond that of the executive functioning tests alone?

1. Method

1.1. Participants

A sample of 46 adults (age: M = 40.87, S.D. = 15.32, range = 19–75 years) were participants in this study. Participants were recruited from consecutive outpatients referred for neuropsychological assessment at Baylor College of Medicine Department of Neurosurgery, Houston, TX (N = 23); Sacred Heart Medical Center Neuropsychology Service, Spokane, WA (N = 18); and Deaconess Medical Center Behavioral Medicine Service, Spokane, WA (N = 5). Diagnoses included epilepsy (34.8%), traumatic brain injury (26.1%), and others (39.1%) (e.g., tumor, vascular accident or malformation, and multiple sclerosis). Fifty-six percent of the sample was male. The average level of education of the sample was 13.48 years (S.D. = 2.50, range = 9–21 years). The average obtained full scale IQ for the sample was 95.91 (S.D. = 14.23, range = 55–127), as measured by the Wechsler Adult Intelligence Scale, 3rd edition (Wechsler, 1997; 59% of the sample) or the Wechsler Abbreviated Scales of Intelligence (Psychological Corporation, 1999; 41% of the sample). Data were collected as part of each participant’s clinically indicated neuropsychological evaluation.

Participants were all native speakers of English, 18 years of age or older, and were able to give consent and understand all test instructions. Time since symptom onset (i.e., when symptoms began, such as the time since head injury or initiation of seizures) was between 3 months and 1 year for 32.6% of the sample, between 1 and 5 years for 23.9% of the sample, and greater than 5 years for 43.5% of the sample. In order to complete all aspects of this study, each participant had to have a significant other who was willing to complete the questionnaires. The majority of significant other informants were immediate family members of the participant (52.2% spouse, 32.6% parent). The remaining significant others comprised other family members or close friends of the participant (15.3%). The significant others were with the participant an average of 82.79 h (S.D. = 59.57, range 5–168) per week, and the majority of the significant others live with the participant (69.6%).

1.2. Materials

1.2.1. Executive functioning tests

The Wisconsin Card Sorting Test (WCST), the Trail Making Test (TMT), the Stroop Color and Word Test (Stroop), and the Controlled Oral Word Association Test (COWAT) were used as neuropsychological tests of executive func-
tioning. These tests were selected because they are commonly used clinical measures that clinicians and researchers typically classify as “executive” although we acknowledge that these tests can also be classified as tests of attention (TMT, Stroop), language (COWAT), and visuomotor speed (TMT) (see Lezak, 1995; Spreen & Strauss, 1998). Further, the executive tests were selected based on their inclusion in previous ecological validity research. A detailed discussion of the complex issues surrounding the conceptualization and measurement of executive functioning is beyond the scope of this paper.

In order to reduce Type I error, only one variable from each measure was selected. Although there are several ways to select the variable of interest from each of the tests, this study took a clinical approach. More specifically, we selected variables based on use in previous ecological validity research and relevance to clinical practice. Given that several of the variables are multifaceted and likely measure abilities beyond just executive functioning, we do not consider the chosen variables to be pure measures of executive functioning, but rather to have a relatively large executive component. Prior research has not typically assessed variable selection issues. Previous studies have examined every variable possible (Burgess et al., 1998; Poole et al., 1999) and/or calculated difference scores that attempt to account for the nonexecutive components of these tests (Burgess et al.; Chan, 2001; Norris & Tate, 2000). Difference scores are not, however, commonly used in a clinical context and they can have psychometric limitations (see Cohen & Cohen, 1983). In addition, the primary goal of this study is not to assess the ecological validity of individual neuropsychological tests, but rather to assess the added contribution of extra-test variables to the group of executive tests together. Therefore, by selecting variables that may have processing speed, verbal, motor, or visuospatial demands in addition to executive functioning, we will likely be accounting for more variance in everyday ability than we would with more pure executive measures. This will then result in a more conservative test of our hypothesis that extra-test factors can add additional variance to prediction of everyday ability.

1.2.1.1. The Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtis, 1993). This test is a standard clinical measure that is commonly used as a measure of executive functioning and is thought to require cognitive flexibility, problem solving, and the use of feedback to guide behavior. In this test, participants are asked to match cards that vary by color, shape, and number to four “key cards.” Participants are not told how to sort the cards, but must determine the correct category from the feedback given by the examiner, which changes periodically throughout the test. The full 128-card version was administered. The percentage of perseverative errors committed during the test was used as the measure of executive functioning from the WCST. This variable was chosen because it has been shown to have sensitivity to frontal lobe lesions and is considered more purely executive than other variables from the WCST (i.e., categories or nonperseverative errors) (Lezak, 1995). This variable has also been used in previous ecological validity research (Burgess et al., 1998; Poole et al., 1999; Ready et al., 2001).

1.2.1.2. The Trail Making Test (Army Individual Test Battery, 1944). This commonly used test consists of two parts. Part A requires the connection by pencil lines of numbers (1–20) positioned randomly on an 8.5 × 11 in. sheet of paper. Part B requires that the participant alternate between numbers and letters in order (e.g., 1–A–2–B–3–C–…). The total time required for completion of both parts was recorded (including any time used by the examiner to point out mistakes and the participant to correct these mistakes). Total time for Trails B was used as the measure of executive functioning from the TMT. Although Trails B involves speeded processing and visual scanning abilities, it has a strong cognitive flexibility component (Korite, Horner, & Windham, 2002, also see Lezak, 1995; Spreen & Strauss, 1998) and has consistently been used in previous ecological validity research (Amieva et al., 2003; Burgess et al., 1998; Chan, 2001; Ready et al., 2001).

1.2.1.3. The Stroop Color and Word Test (Golden, 1978). This test consists of three trials. The first trial requires the participant to quickly read color words. The second trial requires the participant to quickly name the color of “Xs” printed in colored ink, and the third trial requires the participant to name the color of ink that color words are printed in (e.g., the word “blue” is printed in red ink and the participant must say “red”). All trials have a 45 s time limit and errors are corrected. The total score for the color-word (interference) trial was used as the measure of executive functioning from the Stroop (see Lezak, 1995). This variable has also been used in previous ecological validity research (Amieva et al., 2003; Bogod et al., 2003; Chan, 2001).
1.2.1.4. The Controlled Oral Word Association Test (Benton, Hamsher, & Sivan, 1983). This test requires participants to produce as many words as they can that begin with the letters F, A, and S, respectively. The participant is given 60 s for each letter. The total number of words produced across all three trials was used as the measure of executive functioning from the COWAT (see Lezak, 1995).

1.2.2. Questionnaires

The DEX and the BAFQ were administered as measures of everyday executive ability. These questionnaires were completed by significant others. Informant report was used because previous research has suggested that self-report of cognitive ability is only weakly, if at all, related to test performance in neurological populations (Burgess et al., 1998; Evans et al., 1997; Goldstein & McCue, 1995; Kaitano, Koskinen & Kaipo, 1995; Sunderland, Harris, & Baddeley, 1983).

1.2.2.1. The Dysexecutive Questionnaire (Wilson, Alderman, Burgess, Emslie, & Evans, 1996). The DEX (informant report version) required informants to rate, on a Likert-type scale from 0 (never) to 4 (very often), how often they observe each of the 20 executive problems. The internal consistency reliability of this scale was adequate (alpha = .90) and consistent with previous research (alpha = .74). This questionnaire was modified from its original format to allow for two additional questions to be asked after each executive symptom. The first additional question was designed to elicit information about the participant’s usual daily routine and what cognitive skills are typically required of them. For example, for the DEX item “gets events mixed up with each other, and gets confused about the correct order of events”, the informant was also asked to rate on a 5-point Likert Scale from 0 (never) to 4 (very often) how often problems in this area interfere with the participant’s usual daily activities. To obtain a measure of environmental executive demands, this question was asked for each of the 20 executive symptoms from the DEX and the resultant score will be referred to as the “demand score.” The internal consistency reliability of this scale was adequate (alpha = .90). The second additional question (i.e., “how often does he/she do something to compensate for, or prevent, difficulties in this area?”) was designed to elicit information about the compensatory strategies the participant uses to prevent or reduce problems in each symptom area. It was also asked for each of the 20 DEX executive symptoms and ratings were again made on a 5-point Likert scale. This rating was designed as a measure of executive compensatory strategy use and will be referred to as the “strategy score”. The internal consistency reliability of this scale was also adequate (alpha = .86).

Thus, three variables were derived from the same 20 executive problems: (1) the DEX total score (how often each executive problem occurs), (2) the demand score (how much each executive problem interferes with daily life), and (3) the strategy score (how often compensatory strategies are used for each executive problem).

1.2.2.2. The Brock Adaptive Functioning Questionnaire (Dywan & Selandowsky, 1996). The BAFQ informant questionnaire required informants to rate on a Likert-type scale from 0 (hardly ever/never) to 4 (almost always) how often they observe each of the 68 executive problems. This scale is composed of 12 subscales that assess the following executive domains: planning, initiation, flexibility, excess caution, attention, memory, arousal level, emotionality, impulsivity, aggressiveness, social monitoring, and empathy. Although both the DEX and BAFQ assess similar domains of executive functioning, the BAFQ has over three times as many questions as the DEX. This measure was administered to demonstrate the convergent validity of the DEX. The internal consistency reliability of this scale was adequate (alpha = .92).

For both the DEX and BAFQ, the valid responses for each scale were totaled and divided by the number of items answered to derive an average score between 0 and 4. The average scores derived for the DEX and the BAFQ served as the primary outcome measures. This approach was used because replacing missing data with the mean response would have increased the total score, while totaling only valid responses would have treated missing data as a zero, which would both be misleading. For the DEX and BAFQ, missing data accounted for 1.5 and 1.4% of the responses, respectively. The environmental cognitive demand questions contained 3.1% missing data, and the compensatory strategy use questions contained 2.4% missing data.

1.3. Procedure

After a participant was scheduled for a neuropsychological evaluation, and met the study inclusion criteria, informed consent was obtained from both the participant and his/her significant other. The informant questionnaires (DEX and BAFQ) were then filled out and returned at, or prior to, the feedback session. Participants underwent a clinically...
2. Results

The approach to data analysis was as follows. First, to clarify the individual relationships among the executive tests and between the executive tests and measures of everyday executive functioning, correlations were computed among the executive test variables and between the executive test variables and the measures of everyday executive functioning. Second, the executive tests were combined in a hierarchical multiple regression equation predicting each of the two measures of everyday executive functioning (DEX and BAFQ), in order to determine the ecological validity of the executive tests as a whole. Lastly, environmental demand and strategy use were each added to the regression equation (after the group of executive tests) to determine whether these measures improve prediction of everyday executive functioning.

2.1. Correlations

The correlation matrix for the neuropsychological measures of executive functioning was examined. Derived scores from the TMT (B–A) and Stroop (interference score) were also included in the matrix. As can be seen in Table 1, the Stroop Interference score was not significantly related to any other executive measure (with the exception of the Stroop Color–Word Naming trial; Stroop Int., Stroop Interference score; COWAT, Controlled Oral Word Association Test; WCST, Wisconsin Card Sorting Test percent perseverative errors. Scores reflected where appropriate, to make correlations positive.

Note: Trails B, Trail Making Test Part B; Trails B–A, Trail Making Test Part B minus Trail Making Test Part A; Stroop CW, Stroop Color–Word Naming trial; Stroop Int., Stroop Interference score; COWAT, Controlled Oral Word Association Test; WCST, Wisconsin Card Sorting Test percent perseverative errors. Scores reflected where appropriate, to make correlations positive.

* \( p < .05. \)
** \( p < .01. \)

indicated neuropsychological evaluation including intake interview and comprehensive testing, which included the neuropsychological measures described above. Demographic information was obtained during the clinical intake interview. Participants and referring physicians were provided with written and/or oral feedback regarding the participants' cognitive functioning as a routine part of the neuropsychological evaluation.

For interested readers, the correlations between the DEX and Trails B–A \( (r = .34) \) and the Stroop Interference score \( (r = .11) \) were not significant. The correlations between the BAFQ and Trails B–A \( (r = .21) \) and the Stroop Interference score \( (r = .14) \) were also not significant.
and the COWAT were not significantly related to either measure of everyday executive ability. Despite failing to reach statistical significance, the correlations between the COWAT and the outcome measures were in the correct direction and may have reached significance if the sample was larger. There was a near zero correlation, however, between the WCST and both the DEX and BAFQ. The pattern of correlations between the neuropsychological tests and the DEX and the BAFQ were highly similar.

2.2. Hierarchical multiple regression

Hierarchical multiple regression was used to determine the ecological validity of the group of executive tests, and to determine the contribution of extra-test factors (i.e., demand score and strategy score) to prediction of everyday executive functioning. The regression data are presented in Table 3.

2.2.1. Ecological validity

In order to determine the ecological validity of the administered executive functioning tests as a group, the DEX was regressed on Stroop Color–Word score, total words from the COWAT, percent perseverative errors from the WCST, and time on Trails B. The $R^2$ for the entire model was .20, $p = .06$. None of the individual tests were uniquely significant when the other measures were included in the regression equation. This procedure was repeated using the BAFQ as the outcome measure in order to determine if the DEX and BAFQ had similar characteristics and if the model could be replicated using a different measure of everyday executive ability. The $R^2$ for the entire model was .18, $p = .08$. Again, no individual variable uniquely predicted the BAFQ.

2.2.2. Environmental demand and strategy use

The demand score and strategy score were both highly correlated with the DEX (demand, $r = .61$; strategy, $r = .55$) and also with each other ($r = .55$). Since these variables were assessed as a modification to the DEX, it is possible that part of the strong relationships between these variables arose from the use of the same measurement technique and the same reference symptoms, rather than strong relationships between the constructs in question. These strong relationships

Table 3
Hierarchical multiple regression models

<table>
<thead>
<tr>
<th>Block</th>
<th>Dependent variable</th>
<th>$R^2$ change</th>
<th>Overall $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive tests alone</td>
<td>DEX</td>
<td>–</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>BAFQ</td>
<td>–</td>
<td>.18</td>
</tr>
<tr>
<td>Executive tests + Demand Score</td>
<td>BAFQ</td>
<td>25**</td>
<td>.47**</td>
</tr>
<tr>
<td>Executive tests + Strategy Score</td>
<td>BAFQ</td>
<td>15**</td>
<td>.37**</td>
</tr>
<tr>
<td>Executive tests + [Demand + Strategy]</td>
<td>BAFQ</td>
<td>28**</td>
<td>.51**</td>
</tr>
<tr>
<td>Demand Score + Executive tests</td>
<td>BAFQ</td>
<td>16</td>
<td>.47**</td>
</tr>
<tr>
<td>Strategy Score + Executive tests</td>
<td>BAFQ</td>
<td>17</td>
<td>.37**</td>
</tr>
</tbody>
</table>

* The only variable that contributed uniquely to the model ($beta = .46, p < .01$).
* $p < .05$.
** $p < .01$. 

Note: Scores reflected where appropriate to make correlations positive.
* $p < .05$.
** $p < .01$. 

Table 2
Correlations between everyday executive ability and neuropsychological test performance ($N = 46$)

<table>
<thead>
<tr>
<th>Test</th>
<th>DEX</th>
<th>BAFQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trails B</td>
<td>.25</td>
<td>..33</td>
</tr>
<tr>
<td>COWAT</td>
<td>.28</td>
<td>.24</td>
</tr>
<tr>
<td>Stroop Color–Word</td>
<td>.35</td>
<td>.38**</td>
</tr>
<tr>
<td>WCST percent perseverative errors</td>
<td>.05</td>
<td>–.09</td>
</tr>
</tbody>
</table>
make theoretical sense in that people with very demanding environments are more likely to have executive everyday failures that are noticeable to others. Further, people who have significant daily problems with executive functioning are more in need of compensatory strategies to help mitigate these difficulties. To help clarify this issue statistically, when the strategy score was regressed on the DEX and the demand score, the demand score contributed unique variance above and beyond the DEX ($R^2$ change = .07, $p < .05$, beta weight = .34). This suggests that the DEX and the demand score represent two independent constructs. Likewise, the strategy score contributed unique variance above that predicted by the DEX when the demand score was the dependent variable ($R^2$ change = .06, $p < .05$, beta weight = .29), again suggesting that these variables measure different constructs.

In the analysis of environmental cognitive demand and strategy use, only the BAFQ was used as the primary independent variable since the actual test items do not overlap and the BAFQ was completed as a separate questionnaire (demand, $r = .56$; strategy, $r = .45$). This analysis revealed an $R^2$ change of .25 for the addition of the environmental cognitive demand score above and beyond the executive tests, $p < .01$, suggesting that the higher the demand for executive skills in the participant’s everyday environment, the higher the ratings of everyday executive impairment. An $R^2$ change of .15 was obtained for the addition of the strategy score to the group of executive test scores, $p < .01$, suggesting that participants who have more everyday executive problems also use more compensatory strategies in everyday life. To assess if either of these two variables (environmental cognitive demand and compensatory strategy use) contribute unique variance, the demand and strategy scores were both added to the regression equation after the group of executive tests. An $R^2$ change of .28 was obtained, $p < .01$, although only the demand score contributed unique variance to the model ($beta = .46$, $p < .01$).

Another related question was addressed next. It is possible that accounting for variance in environmental cognitive demand and compensatory strategy use could clarify the relationship between the executive tests and the measures of everyday executive ability. Therefore, the BAFQ was regressed on the demand score. Next, the group of executive tests were added to the equation, resulting in an $R^2$ change of .16, $p < .05$. Thus, the group of executive tests was significantly related to the BAFQ after the demand variance was removed. When the strategy score was entered into the regression before the group of executive tests, the $R^2$ change approached significance ($R^2$ change = .17, $p = .07$). Given the comparable amount of variance accounted for, the failure to attain significance is likely due to diminished power for the latter analysis, as the sample size was slightly smaller for this analysis (two fewer participants completed the strategy use questions than the demand questions).

3. Discussion

The main objective of the current study was to investigate ways to improve the ecological validity of the neuropsychological assessment of executive functioning. This study demonstrated empirically that assessment of nontraditional variables, such as compensatory strategy use and environmental cognitive demands, can account for additional variance in informant ratings of everyday executive functioning beyond that accounted for by traditional neuropsychological tests of executive functioning. Additionally, accounting for these variables can improve the relationships between the executive tests and informant ratings of everyday executive ability. Although several authors had speculated that these variables are important to understanding ecological validity, empirical research was lacking. The relationship between performance on neuropsychological tests and everyday cognitive ability is complex and multifaceted and research needs to reflect this complexity.

The correlations between the individual executive test variables and the questionnaires obtained in this study were generally lower than those reported in several studies that have used the informant report version of the DEX and similar measures of executive functioning (e.g., Burgess et al., 1998). Other studies, however, have reported null findings using similar methodology, including the informant version of the DEX and similar neuropsychological measures of executive functioning (Bogod et al., 2003; Chan, 2001; Evans et al., 1997; Norris & Tate, 2000). Thus, based on this study and previous research, adequate ecological validity of individual neuropsychological tests is not universal, and may vary significantly by the particular test, even when using the same outcome measure. In this study, Trails B and the Stroop Color–Word score were found to have the highest ecological validity, while the WCST perseverative errors was not related to everyday executive functioning. The COWAT, although not significant, was marginally related to everyday executive functioning. It should be noted that both the Trails B and Stroop Color–Word score involve speeded processing in addition to executive functioning, which may partially account for the higher ecological validity of these two tasks. Thus, it is possible that the significant relationships between these two measures
and the measures of everyday executive functioning may not be due entirely to the executive function component of these tasks.

Since most clinical evaluations involve the use of several measures of executive functioning, looking at individual correlations may be somewhat misleading. Therefore, the tests were analyzed as a block when predicting everyday executive skill. Although not statistically significant, the block of executive tests did account for a clinically meaningful amount of variance in everyday executive ability (18–20%). But this relationship was far from perfect, with as much as 80% of the variance unaccounted for.

As hypothesized, inclusion of a measure of environmental cognitive demand accounted for significantly more variance in everyday executive functioning than the executive tests alone. Likewise, adding the measure of compensatory strategy use also accounted for significantly more variance than the executive measures alone. This supports the hypothesis that these variables are important to assess when trying to predict everyday executive ability. Furthermore, accounting for differences in environmental demand and compensatory strategy use improved the relationship between the executive tests and the outcome measure. When controlling for environmental cognitive demands, the executive tests were significant as a block. When controlling for compensatory strategy use, the executive tests approached significance. Thus, by controlling for the variance accounted for by these variables, the ecological validity of the group of executive tests was improved. It appears that differences in these variables can obscure the relationship between neuropsychological tests and everyday ability, as suggested by several writers. The current study provides initial evidence in support of this common belief, although more research is needed. More research is also needed to explore the mechanisms behind these relationships and which variables are most important to assess. The current study provides preliminary data to suggest that environmental cognitive demand may play a larger role than compensatory strategy use, as it provided unique variance in the prediction of everyday executive functioning.

There are several limitations to the current study. First, the sample size is relatively small, which resulted in keeping the number of tests examined to a minimum. Consequently, only four tests of executive functioning were included in the current study and it is possible that the results would have been different if a different set of tests, or variables from these tests, were chosen. However, the executive tests evaluated in this study are commonly used in clinical practice and were thought to be of the most clinical relevance. In addition, because the primary goal of the current study was to determine the amount of variance added by extra-test variables above and beyond the executive tests, the fact that some of the measures also included nonexecutive components (e.g., speeded processing, visual scanning) would actually reduce the likelihood of accounting for unique variance. In general, variable and test selection is an important issue that has not been addressed well in the ecological validity literature. Future research should address this issue.

The choice of criterion measure is an inherent limitation of all research on the ecological validity of neuropsychological tests, as one can never really know the level of true everyday ability. Thus, two different measures of everyday executive ability were administered to determine if similar results would be obtained. In this study, ecological validity was measured against a significant other rating of executive ability in everyday life. There are advantages and limitations of this approach. The literature suggests that self-report is a weaker measure of everyday cognitive performance than clinician and informant ratings in neurologically impaired individuals (Burgess et al., 1998; Chaytor & Schmitter-Edgecombe, 2003; Evans et al., 1997; Goldstein & McCue, 1995; Kaitaro, Koskinen, & Kaipio, 1995; Sunderland, Harris, & Baddeley, 1983); therefore, significant other ratings were employed in this study. However, there is unavoidable error involved in this approach as well. The ratings are only as good as the person performing them. Future research should explore the convergent validity across different methods of assessing everyday ability, such as simulations, clinician ratings, and significant other ratings. We chose a questionnaire-based rating for logistical purposes, as it was the most time-efficient and practical way of assessing ecological validity. We understand that this is only an estimate of true everyday executive ability.

A review of the literature on the ecological validity of neuropsychological tests pointed out several variables that could possibly affect ecological validity research (Chaytor & Schmitter-Edgecombe, 2003). One such variable is population effects, such that neuropsychological tests may have different degrees of ecological validity in different populations. Since the current study employed a general neurological sample to most closely approximate the typical clinical practice, it is possible that population effects may have obscured some findings. To explore this possibility, the initial set of correlations between the executive tests and the DEX and BAFQ were run separately for the participants with epilepsy (N=16) and those with traumatic brain injury (TBI, N=12). Given the small number of participants in each group, these results are only provided to observe the overall trend, not to examine any individual statistic or significance levels. Examination of the separate correlation matrices compared to the results for the entire sample,
revealed some striking differences. First, for the group of TBI patients, the correlations between the COWAT and both the DEX (r = .50), and BAFQ (r = .31) were substantially larger than in the overall sample (DEX, r = .28; BAFQ, r = .23). For the group of epilepsy patients, the correlations between the WCST and both the DEX (r = .30), and BAFQ (r = .24) were substantially larger than in the overall sample (DEX, r = .03; BAFQ, r = -0.8). Additionally, for the epilepsy group, the correlations between Trails B and both the DEX (r = .50), and BAFQ (r = .54) were substantially larger than in the overall sample (DEX, r = .25; BAFQ, r = .33). These findings provide some interesting preliminary evidence to suggest that the ecological validity of the executive measures may vary across populations. The epilepsy group in this sample had predominantly temporal lobe pathology, while the TBI group would be expected to have more frontal lobe pathology. In the TBI sample, the COWAT may be operating as a more frontal measure, while in the temporal lobe epilepsy group it may be assessing language ability and therefore may not be as highly related to an executive outcome measure. This exploratory finding suggests that future research on ecological validity should more systematically examine population effects.

4. Conclusion

In summary, the current study adds to the literature on the ecological validity of executive functioning assessment by highlighting the multitude of factors that are important to assess when trying to understand the complex relationship between cognitive testing and real world performance. First, among the executive tests administered in this study, two out of four were not significantly related to either measure of everyday executive functioning, and the tests that were significantly related to everyday executive functioning, may have been related because of nonexecutive components. This is consistent with previous research in suggesting that not all executive tests have adequate ecological validity. As a whole, the current set of commonly used executive tests failed to significantly predict everyday executive functioning. The executive tests combined accounted for 18–20% of the variance in everyday executive ability. Adding assessment of environmental cognitive demand and compensatory strategy use to the set of executive tests resulted in a model that accounted for 51% of the variance in everyday executive ability. As several theorists have suggested, environmental cognitive demands and compensatory strategy use do appear to affect the ability of neuropsychological tests to predict real world behavior and should be explored in future research on ecological validity. In order to improve the ecological validity of executive functioning assessment, research needs to go beyond the tests themselves and attempt to empirically investigate these complex relationships.

Acknowledgement

This project was completed in partial fulfillment of the first author’s doctoral degree in clinical psychology at Washington State University. We thank Michelle Kibby and Craig Parks for their help and insightful comments. We gratefully acknowledge Harvey Levin and Michele York for allowing data collection at Baylor College of Medicine, and Michelle White for allowing data collection at Deaconess Medical Center. Heather Nissley, Matthew Wright and Ellen Woo are also recognized for their assistance with data collection.

References


