IMPROVING THE ECOLOGICICAL VALIDITY OF

EXECUTIVE FUNCTIONING ASSESSMENT

By

NAOMI SAGE CHAYTOR

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
Department of Psychology

AUGUST 2004

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of NAOMI SAGE CHAYTOR find it satisfactory and recommend that it be accepted.

____

Chair

ACKNOWLEDGEMENT

I would like to acknowledge the guidance and support of my mentor, Dr. Maureen Schmitter-Edgecombe. Thank you for your dedication to this research project, as well as to my professional development. You have supported my research ideas and clinical interests throughout my training. I would also like to acknowledge Dr. Robert Burr for allowing me to collect data on his patients and for his practical support throughout this process. And to my committee, thank you for your helpful assistance and guidance with this project. I would also like to thank Dr. Michele York and Dr. Harvey Levin for their professional guidance and for allowing me to collect data at Baylor. And finally, I would like to thank Michael McDonell for his unconditional support of everything I do.

IMPROVING THE ECOLOGICAL VALIDITY OF

EXECUTIVE FUNCTIONING ASSESSMENT

Abstract

by Naomi Sage Chaytor, Ph.D.
Washington State University
August 2004

Chair: Maureen Schmitter-Edgecombe

Although studies have found significant relationships between executive functioning tests and measures of everyday cognitive skills, the magnitude of the relationships tend to be in the moderate range at best. The main objective of the current study was to investigate ways to improve the ecological validity of the neuropsychological assessment of executive functioning through the formal assessment of compensatory strategies and environmental cognitive demands. Factor analysis of the measures of everyday executive functioning was also conducted in order to determine if more specific relationships between tests and outcome could be identified. Results indicated that the group of executive functioning tests together (i.e., Trail Making Test, Wisconsin Card Sorting Test, Stroop Color-Word Test, and Controlled Oral Word Association Test) accounted for 18-20 percent of the variance in everyday executive ability. The addition of measures of environmental cognitive demand and compensatory strategy use each significantly increased the variance in the everyday executive ability accounted for. In addition, at least three reliable executive factors were identified in this and other studies (i.e., Inhibition, Intentionality and Executive Memory). 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.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

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.

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, Emslie & 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). Several other studies investigating the ecological validity of executive functioning tests have also found significant relationships between neuropsychological tests and outcome measures (Clark et al., 2000;

Dimitrov et al., 1996; Evans et al., 1997; Poole et al., 1999; 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 (e.g., DEX) have found very little relationship between tests of executive functioning and measures of everyday cognitive skills (Amieva, Phillips & Della Sala, 2003; Bogod et al., 2003; Chan, 2001; Norris & Tate, 2000; Ready et al., 2000). Currently, however, it remains difficult to compare findings across studies as the studies have differed 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; Franzen & 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 non-cognitive factors that can influence the relationship between test performance and everyday performance. These non-cognitive 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 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 the 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.

It is also possible that meaningful relationships between neuropsychological tests of executive functioning and everyday executive ability are being obscured because executive functioning is not a unitary construct. Thus, a given test may be related to one aspect of executive functioning but not another. If only an overall score is used as the outcome measure, these more specific relationships may be missed. To address this issue, Burgess et al. (1998) investigated the factor structure of the DEX, and found a 5-factor solution. This suggests that the dysexecutive syndrome is multidimensional at the behavioral level. The authors labeled the five factors Inhibition, Intentionality, Executive Memory, Positive Affect, and Negative Affect. Several neuropsychological measures were related to the first three factors, while none of the neuropsychological tests were related to the positive and negative affect factors.

Two other factor analytic studies on the DEX also found a five-factor solution. In a study by Amieva et al. (2003), the following factors were identified: intentionality, interference management, inhibition, planning, and social regulation. The inhibition factor was found to be related to the Stroop task. This study had a very small sample size (N = 20) and the results should, therefore, be interpreted with caution. Chan (2001) investigated the factor structure of the DEX in a larger sample of normal Chinese participants (*N* = 93). The factors were labeled inhibition, intentionality, knowing-doing dissociation, in-resistance, and social regulation. All five factors were correlated with the Trail Making Test, while the inhibition factor was also correlated with the Stroop. As a whole, these studies have consistently found factors corresponding to inhibitory processing, planning or intentionality, and executive memory (although the labels for these factors have varied somewhat across studies), while the remaining two factors have not been consistently reported.

Since all the factor analytic studies described above have used the DEX as the measure of everyday executive functioning, it is possible that the similar factor solutions are due to mono-method bias. Using the Brock Adaptive Functioning Questionnaire (BAFQ), another measure of everyday executive functioning designed to assess the dysexecutive syndrome in traumatic brain injured patients, Simpson and Schmitter-Edgecombe (2002) found a two factor solution. These two factors were labeled dorsolateral and orbitofrontal. The functions assessed by the first factor have been associated in the literature with functioning in the dorsolateral area and likewise for the orbitofrontal factor. This work was partially based on another factor analytic study on 199 normal adolescents using the BAFQ (Dywan, Roden & Murphy, 1995), which also found a two factor solution. Thus, these studies also found evidence for the fractionation (although with fewer factors) of the dysexecutive syndrome, while using a different measure of everyday executive functioning.

This body of research raises an important issue, as it suggests that executive dysfunction is not a unitary construct. The factor analytic studies to date seem to suggest that inhibition, intentionality or planning, and executive memory are important distinct elements of the dysexecutive syndrome as measured by the DEX. Beyond these three factors, there is disagreement in the literature, as factors labeled positive and negative affect, in-resistance, knowing-doing dissociation, and social regulation have all been identified as well. Further clarification of the true factors underlying the dysexecutive syndrome is needed. Most studies have found that a subset of the factors were not related to any of the neuropsychological tests administered. This indicates that there may be aspects of the dysexecutive syndrome that are not adequately sampled by typical neuropsychological batteries. Additionally, the studies by Simpson & Schmitter-Edgecombe (2002) and Dywan et al. (1995) found a different factor solution when using the BAFQ. Replication of meaningful factors across measures is important to ensure that the underlying factors are not just specific to one measure.

The main objective of the current study is to 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 everyday executive skills beyond that accounted for by traditional neuropsychological tests of executive functioning. Additional goals for the current study included replication of previous work investigating the ecological validity of neuropsychological tests of executive functioning, as well as attempting to replicate the fractionation of the dysexecutive syndrome. Because a different number of factors have been identified when using the DEX versus the BAFQ, an additional goal of the current study was to attempt to reconcile these findings in one study. The following are the specific research questions that were addressed in the current study.

1. Are the executive functioning tests related to everyday executive functioning skills, and are these relationships greater than those with non-executive tests?

2. Does assessment of environmental cognitive demands account for additional variance in everyday executive functioning beyond executive functioning tests?
3. Does assessment of compensatory strategies account for additional variance in everyday executive functioning beyond executive functioning tests?
4. Can previous work on the fractionation of the dysexecutive syndrome be replicated using similar outcome measures and neuropsychological tests?

CHAPTER TWO

METHOD

Participants

A sample of 46 adults (age; *M* = 40.87, *SD* = 15.32, *range* = 19-75 years) were participants in this study. Participants were recruited from consecutive out-patients 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 other (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 (*SD* = 2.50, *range* = 9-21 years). The average obtained full scale IQ for the sample was 95.91 (*SD* = 14.23), as measured by either the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III; 59%), or the Wechsler Abbreviated Scales of Intelligence (WASI; 41%). No significant difference in full scale IQ was found between those participants receiving the WASI (*M* = 99.00, *SD* = 10.18) versus the WAIS-III (*M* = 93.74, *SD* = 16.34), $t(44) = 1.34$, $p > .05$. 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 one year for 32.6 percent of the sample, between 1 and 5 years for 23.9 percent of the sample, and greater than 5 years for 43.5 percent 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 (see method section).

Materials

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 functioning. 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 applicability 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. In addition to the variables that will be used in this study, previous studies have presented every variable possible (Burgess et al., 1998; Poole et al., 1999), and/or calculated difference scores that attempt to account for the non-executive components of these tests (Burgess et al.; Chan; Norris & Tate, 2000). Difference scores are not commonly used in a clinical context and they can have psychometric limitations, therefore, we chose not to use difference scores. In addition, the primary goal of this study is not to determine the ecological validity of individual neuropsychological tests, but to assess the added contribution of extra-test variables. 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. Thus, this will allow for a more conservative test of our hypothesis that extra-test factors can add additional variance to prediction of everyday ability.

The Wisconsin Card Sorting Test (Heaton, 1981). This test is a standard clinical measure that loads heavily on executive functioning and requires 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 used. 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 non-perseverative errors) (Lezak, 1995). This variable has also been used in previous ecological validity research (Burgess et al., 1998; Pool et al., 1999; Ready et al., 2001).

The Trail Making Test (Reitan & Wolfson, 1995). 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 ½ x 11-inch 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. Trails B appears to be a measure of cognitive flexibility (Kortte, Horner & Windham, 2002, also see Lezak, 1995, Spreen & Strauss 1998) and has been used in previous ecological validity research (Amieva et al., 2003; Burgess et al., 1998; Chan, 2001; Ready et al., 2001). Although a Trails B – Trails A computation has been used in previous studies (Burgess et al.; Chan; Norris & Tate, 2000), there are psychometric problems with this type of procedure and it was therefore not used in this study.

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 second 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). Although an "interference score" that controls for reading and color naming speed can be calculated, it was not used in the current study because the resulting variable tends to have a restricted range.

The Controlled Oral Word Association Test (Spreen & Benton, 1977). 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 seconds 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).

Other Neuropsychological Tests

In order to establish the discriminant validity of the executive functioning tests, several tests of neuropsychological functioning were also administered with the aim of covering the broad areas of general intellectual functioning, language, and memory.

Wechsler Adult Intelligence Scales-Third Revision, Verbal IQ (WAIS-III; Wechsler, 1997) or *Wechsler Abbreviated Scales of Intelligence, Verbal IQ* (WASI; Psychological Corporation, 1999). These are widely used measures of general intellectual abilities that can be broken down into a Verbal and Performance IQ score. The Verbal IQ score was used because several Performance subtests involve executive functioning abilities (i.e. Matrix Reasoning).

Boston Naming Test (BNT; Kaplan, Goodglass & Weintraub, 1978). This is a test of visual confrontation naming. The participant must identify each of 60 line drawings. The total number correct was used as a measure of language ability.

California Verbal Learning Test-II (CVLT-II: Delis, Kramer, Kaplan & Ober, 1987). This is a test of verbal list-learning and recall. The participant is read a list of 16 items and asked to recall as many items as possible. This procedure is repeated 4 more times. A distractor list is then administered, which is followed by free recall and then semantically cued recall of the original list. After a 20-minute delay, free and cued recall of the list is assessed. This is followed by a recognition trial. The total number of items correctly recalled during the long-delay free recall trial was used as a measure of verbal memory.

Questionnaires

The DEX and the BAFQ were administered. These questionnaires were completed by significant others. Informant report was used because 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; Kaitaro, Koskinen & Kaipio, 1995; Sunderland, Harris & Baddeley, 1983).

The Dysexecutive Questionnaire (Wilson, Alderman, Burgess, Emslie & Evans, 1996). The DEX (informant report version) required informants to rate, on a Likert-type scale from zero (never) to 4 (very often), how often they observe each of 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 symptom (see Appendix A). The first additional question was designed to illicit information about the participant's usual daily routine and what cognitive skills are typically required. 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 asked "how often do problems in this area interfere with his/her usual daily activities?" This rating was designed as a measure of environmental executive demands and 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 illicit information about the compensatory strategies the participant uses to prevent or reduce problems in each symptom area. 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).

The Brock Adaptive Functioning Questionnaire (Dywan & Segalowitz, 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 68 executive problems (see Appendix B). This scale is divided into 12 subscales assessing 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 percent of the responses, respectively. The environmental cognitive demand questionnaire contained 3.1 percent missing data, and the compensatory strategy questionnaire contained 2.4 percent missing data. For the BAFQ, the 12 subscales were also calculated as average scores (no individual scale with more than 2.7 percent missing data).

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 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.

CHAPTER THREE

RESULTS

Correlations

In order to examine the ecological validity of the executive tests, the individual executive (WCST, TMT, Stroop, COWAT) and non-executive (BNT, VIQ, and CVLT-II) variables were correlated with the DEX and BAFQ scores, as can be seen in Table 1. There was a high degree of relatedness between the two measures of everyday executive functioning ($r = .84$, $p <.001$), suggesting that these two scales are measuring the same basic construct. As expected, the BNT and CVLT-II long delay free recall were not related to everyday executive functioning ability (i.e., DEX and BAFQ), although VIQ was significantly correlated with both the DEX ($r = .45$, $p < .01$) and BAFQ ($r = .33$, $p < .05$). The Stroop color-word score was the only executive measure significantly correlated with the DEX ($r = .35$, $p < .05$). Significant correlations were observed between the BAFQ and both the Stroop color-word score ($r = .38$, $p < .01$) and Trails B ($r = .33$, $p < .05$)[1]. The WCST perseverative errors and the COWAT were unrelated 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.

[1] The correlations between the DEX and Trails B-A ($r = .14$) 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.

Table 1. Correlations between everyday executive ability and neuropsychological test performance (*N* = 46).

Test	DEX	BAFQ
Executive tests		
Trails B	.25	**.33***
COWAT	.28	.24
Stroop Color-Word	**.35***	**.38****
WCST % perseverative errors	.03	-.09
General cognitive functions		
VIQ	**.45****	**.33***
BNT	.20	.18
CVLT-II Long delay free recall	.17	.25

Note. Scores reflected where appropriate. *p <.05, **p <.01

Hierarchical multiple regression

Hierarchical multiple regression was used to determine the ecological validity and discriminant 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 2.

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 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.

Table 2. Hierarchical multiple regression models.

Block	Dependent variable	*R^2 change*	*overall R^2*
Executive tests alone	DEX	--	.20
	BAFQ	--	.18
Non-executive tests + Executive tests	DEX	.11	**.32***
	BAFQ	.09	.23
Executive tests + Non-executive tests	DEX	.11	**.32***
	BAFQ	.05	.23
Executive tests + Demand Score	BAFQ	**.25****	**.47****
Executive tests + Strategy Score	BAFQ	**.15****	**.37****
Executive tests + [Demand[1] + Strategy]	BAFQ	**.28****	**.51****
Demand Score + Executive tests	BAFQ	**.16***	**.47****
Strategy Score + Executive tests	BAFQ	.17	**.37****

[1]The only variable that contributed uniquely to the model (*Beta* = .46, *p* <.01).
**p* <.05, ***p* <.01.

Discriminant validity. In order to determine the discriminant validity of the executive functioning measures, the DEX score was regressed on the set of non-executive tests (CVLT long-delay free recall, BNT and VIQ). These variables accounted for 21 percent of the variance in everyday executive functioning as assessed by the DEX, *p* <.05. The set of executive tests were then added to the regression equation. The R^2 change for the addition of this latter set of variables was .11, *p* >.05, indicating that the executive tests did not contribute significantly unique variance in predicting everyday executive functioning above and beyond the general neuropsychological tests. This analysis was repeated using the BAFQ as the dependent variable. The results were similar to those obtained using the DEX, as the executive tests failed to contribute unique variance above and beyond the non-executive tests, R^2 *change* = .09, p >.05, although the set of non-executive tests alone failed to significantly predict the BAFQ, R^2 = .14, p >.05. When this analysis was re-run entering the blocks of variables in the reverse order, the non-executive tests also failed to contribute unique variance beyond that predicted from the executive tests (DEX: R^2 *change* = .11, p >.05; BAFQ: R^2 *change* = .05, p >.05). It should be noted, however, that the number of variables included in this analysis (seven variables) exceeds the power of the current study to detect a medium effect size.

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 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 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, only the BAFQ was used as the primary dependent 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 demand score, 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 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 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 questionnaire than the demand questionnaire).

Factor Analysis

Because previous research has suggested that the dysexecutive syndrome is fractionable, using the overall DEX and BAFQ scores may obscure more specific relationships between subscales and the executive tests. Therefore factor analysis of the two outcome measures was conducted.

DEX. A principle components factor analysis was conducted on the DEX in order to verify the presence of five factors. Principle components extraction with varimax rotation was performed on the responses for the 20 items. Five factors were derived, each with an eigenvalue greater than one, accounting for 71.85 % of the variance in the DEX. The variance accounted for by each factor was as follows: factor one 18 %, factor two 17 %, factor three 15 %, factor four 12 % and factor five 10 %. Factors 1, 2, 3, and 5 had satisfactory reliabilities (alpha = .83, .84, .75, and .72, respectively), while the reliability of factor 4 was lower, alpha = .56.

The pattern of factor loadings obtained on the rotated component matrix was then examined and items with factor loadings greater than .50 were considered significant (Burgess et al., 1998). The factor loadings are presented in Table 3. Table 3 also shows that there is considerable overlap between the factor solutions reported previously for the DEX and the present factor solution. More specifically, factors one, two and three correspond well to the inhibition, intentionality, and executive memory factors reported in previous research (Amieva et al., 2003; Burgess et al.; Chan, 2001). However, factor four was similar to only one previous factor solution (Chan), and factor five was not consistent with any previous reported factor solutions using the DEX. Factor four contains items having to do with awareness of deficits, such as lack of insight, confabulation and saying one thing but doing another, while factor five consists of items suggestive of agitation or hyperactivity.

Table 3. Factor structure of the DEX.

Behavioral Characteristic	Factor loading
Factor 1: Inhibition (alpha = .83)	
13. Disregard for social norms [1, 3]	.86
20. Disregard for how others' feel about his/her behavior [1]	.78
2. Impulsivity [1]	.69
12. Loses temper easily	.66
9. Disinhibition (does embarrassing things) [1, 3]	.62
11. Difficulty showing emotion	.57
Factor 2: Goal-Directed Behavior/Intentionality (alpha = .84)	
19. Poor decision making [1, 2, 3]	.83
10. Variable motivation	.73
8. Apathy/lack of drive	.66
4. Planning problems [1, 2, 3]	.61
5. Euphoria	.59
Factor 3: Executive Memory/Cognition (alpha = .75)	
6. Temporal sequencing problems [1, 2]	.81
1. Impaired abstract verbal comprehension [2, 3]	.80
18. Distractibility [2]	.73
14. Perseveration [1, 3]	.53
Factor 4: Lack of Awareness (alpha = .56)	
7. Lack of insight [3]	.79
3. Confabulation	.69
17. Knowing-doing dissociation [3]	.52
Factor 5: Agitation/Hyperactivity (alpha = .72)	
15. Restlessness	.77
16. Response suppression problems	.71

Note. Only factor loadings over .5 are included. [1]Item loaded on the corresponding factor in the Burgess et al. (1998) study. [2]Item loaded on the corresponding factor in the Amieva et al. (2003) study. [3]Item loaded on the corresponding factor in the Chan (2001) study.

Each factor score, calculated by deriving the average score for the items that loaded on each factor, was then correlated with the executive and non-executive test variables (see Table 4). Similar to the results derived for the overall DEX score, the COWAT and WCST perseverative error scores did not correlate significantly with any of the five factor scores, although the correlation between the COWAT and factor three (executive memory) approached significance ($r = .28$, $p = .058$). The correlations between the WCST and the factor scores were near zero or in the reverse direction. Trails B and the Stroop color-word score were correlated with both the executive memory (Trails B, $r = .35$, $p < .05$; Stroop, $r = .41$, $p < .01$) and awareness (Trails B, $r = .30$, $p < .05$; Stroop, $r = .35$, $p < .05$) factors. The BNT also correlated with the awareness factor ($r = .34$, $p < .05$). The VIQ correlated with all factors except factor 5 (agitation/hyperactivity), and Factor 5 (agitation/hyperactivity) was not correlated with any of the other neuropsychological measures. The majority of the significant correlations were between the neuropsychological measures and the executive memory/cognition factor and the awareness factor. All the correlations reported below were also calculated using regression-based factor scores in order to determine if this would provide a more pure measure of each factor. The correlations were highly consistent with those reported, with the exception that the CVLT-II was significantly correlated with the executive memory factor ($r = .35$, $p < .05$).

Table 4. Correlations between the DEX factor scores and neuropsychological tests.

Test	1	2	3	4	5
Executive tests					
Trails B	.07	.11	**.35***	**.30***	.05
COWAT	.16	.22	.28	.20	.10
Stroop	.18	.21	**.41****	**.35***	.13
WCST	-.01	-.19	.05	.11	-.07
General cognitive functions					
VIQ	**.43****	**.30***	**.42***	**.30***	.18
BNT	.22	-.03	.13	**.34***	.19
CVLT-II	.10	.08	.27[a]	.08	.02

Note. 1 = Inhibition, 2 = Goal-Directed Behavior/Intentionality, 3 = Executive Memory/Cognition, 4 = Awareness, 5 = Agitation/Hyperactivity. For clarity, scores are reflected where appropriate to make correlations positive.

[a] Correlation was significant when using regression based factor scores.

**p* <.05, ***p* <.01.

Since the DEX appears to be assessing several distinct domains of executive functioning, it is possible that the executive tests as a group may be more related to some factors than to others. Therefore, each of the factor scores was regressed on the group of executive tests separately. The set of executive tests accounted for 23 percent of the variance in the executive memory/cognition factor, $p < .05$. The executive tests failed to significantly predict the other four factor scores (Inhibition, $R^2 = .05$; Intentionality, $R^2 = .18$; Awareness, $R^2 = .15$; Agitation/hyperactivity, $R^2 = .05$). The demand and strategy use variables were not included here because of the similar format of these measures and the DEX.

BAFQ. A principle components factor analysis was conducted on the BAFQ in order to verify the presence of two factors (Simpson & Schmitter-Edgecombe, 2002). Principle components extraction with varimax rotation was performed on the mean responses for the 12 subscales. Four factors were derived, each with an eigenvalue greater than one, accounting for 73.01 % of the variance in the BAFQ. The variance accounted for by each factor was as follows: factor one 26 %, factor two 17 %, factor three 17 %, and factor four 13 %. Factors 1, 3, and 4 had satisfactory reliabilities (alpha = .80, .79, and .73, respectively), while the reliability of factor 2 was lower, alpha = .65. The pattern of factor loadings obtained on the rotated component matrix was examined and subscales with factor loadings greater than .5 were considered significant. The factor loadings are presented in Table 5. As can be seen in Table 5, the first three factors correspond well to the first three factors from the factor analysis of the DEX, while the fourth factor did not, although only one subscale loaded highly on this factor (empathy).

Table 5. Factor structure of the BAFQ.

Subscale	Factor loading
Factor 1: Inhibition (alpha = .80)	
Impulsivity	.85
Emotionality	.78
Aggressiveness	.68
Initiation	.62
Factor 2: Goal-Directed Behavior/Intentionality (alpha = .65)	
Excess Caution/Rigidity	.84
Flexibility	.71
Arousal Level/Apathy	.64
Factor 3: Executive Memory/Cognition (alpha = .79)	
Social Monitoring	.75
Memory	.72
Attention	.65
Planning	.52
Factor 4: Empathy (alpha = .73)	
Empathy	.87
Planning	.54

Note. Only factor loadings over .5 are included.

Each factor score from the BAFQ, calculated by deriving the average score for the items that loaded on each factor, was then correlated with the executive and non-executive test variables (see Table 6). Similar to the results derived from the overall BAFQ score, the COWAT and WCST were not significantly related to any of the four factor scores, although the correlation between the WCST and factor three (executive memory) was significant when using the regression-based factor scores ($r = .30$, $p < .05$). Trails B ($r = .34$, $p < .05$) and the Stroop color-word score ($r = .40$, $p < .01$) were both correlated with the executive memory factor, while the Stroop was also correlated with the inhibition factor ($r = .33$, $p < .05$). VIQ was correlated with the inhibition ($r = .36$, $p < .05$) and intentionality factors ($r = .33$, $p < .05$), while the CVLT-II was related to the executive memory factor ($r = .36$, $p < .05$) and the BNT was related to the inhibition factor ($r = .30$, $p < .05$). The majority of the significant correlations were between the neuropsychological measures and the executive memory/cognition factor and the inhibition factor. Factor four (empathy) was not related to any neuropsychological measure. Given the pattern of correlations, the BAFQ factors appear to be measuring generally similar cognitive functions as those measured by the DEX factors, although the pattern was slightly different. All of the correlations reported here were highly consistent with those derived using regression-based factor scores, with the exception of a significant correlation between the WCST and the executive memory factor as discussed above.

Table 6. Correlations between the BAFQ factor scores and the neuropsychological tests.

Test	1	2	3	4
Executive tests				
Trails B	.29	.27	**.34***	.06
COWAT	.26	.18	.18	.10
Stroop	**.33***	.25	**.40****	.23
WCST	.06	-.06	.23[a]	-.01
General cognitive functions				
VIQ	**.36***	**.33***	.14	.17
BNT	**.30***	.08	.11	.04
CVLT-II	.12	.28	**.36***	.04

Note. 1 = Inhibition, 2 = Goal-Directed Behavior/Intentionality, 3 = Executive Memory/Cognition, 4 = Empathy. For clarity, scores are reflected where appropriate to make correlations positive.

[a] Correlation was significant when using regression based factor scores.

*p <.05, **p <.01.

Since the BAFQ appears to be assessing several distinct domains of executive functioning, it is possible that the executive tests as a group may be more related to some factors than to others. Therefore, each of the factor scores was regressed on the group of executive tests separately (see Table 7). Although the following set of analyses should be considered exploratory given the large number of separate regression analyses conducted, we hoped to determine if there was a consistent pattern of findings, where some facets of the dysexecutive syndrome are more related to some variables and not others. Therefore, these analyses were evaluated as a whole rather than interpreting any single finding. The set of executive tests failed to significantly predict any of the four factor scores (Inhibition, R^2 = .15; Intentionality, R^2 = .13; Executive memory/cognition, R^2 = .19; Empathy, R^2 = .06). To determine if the demand score was related to some factor scores and not others, it was added to the regression equation after the executive tests for each factor score. The demand score added significantly to the model for all four factor scores (R^2 change = .25, .13, .15, and .13, respectively). To determine if the strategy score was related to some factor scores and not others, it was added to the regression equation after the executive tests for each factor score. The strategy score added significantly to the model for only the intentionality and executive memory/cognition factors (R^2 change = .13 and .15, respectively). Adding both the demand and strategy score after the executive tests resulted in a significant R^2 change for all factors except the empathy factor.

Table 7. Amount of variance in each BAFQ factor accounted for by the neuropsychological tests of executive functioning and extra-test factors.

Factor	R^2 *change*	*Overall* R^2
Factor 1: Inhibition		
Executive tests	--	.15
Executive tests + Demand	**.25****	**.48****
Executive tests + Strategy	.07	**.32***
Executive tests + Demand + Strategy	**.26****	**.50****
Factor 2: Goal-Directed Behavior/Intentionality		
Executive tests	--	.13
Executive tests + Demand	**.13***	**.27***
Executive tests + Strategy	**.13***	**.27***
Executive tests + Demand + Strategy	**.18***	**.32***
Factor 3: Executive Memory/Cognition		
Executive tests	--	.19
Executive tests + Demand	**.15****	**.35****
Executive tests + Strategy	**.15****	**.38****
Executive tests + Demand + Strategy	**.19****	**.42****
Factor 4: Empathy		
Executive tests	--	.06
Executive tests + Demand	**.13***	.21
Executive tests + Strategy	.01	.14
Executive tests + Demand + Strategy	.13	.26

**p* <.05, ***p* <.01.

The following analyses were conducted in order to determine if extra-test factors moderate the relationship between the executive tests and the factor scores. Accounting for environmental cognitive demand resulted in a significant R^2 change for the set of executive tests only when predicting the inhibition (R^2 change = .15) and executive memory (R^2 change = .21) factors from the BAFQ. Accounting for compensatory strategy use resulted in a significant R^2

change for the set of executive tests only when predicting the executive memory factor from the BAFQ (R^2 change = .25). Accounting for both variables resulted in a significant R^2 change for the set of executive measures only when predicting the executive memory factor (R^2 change = .25).

CHAPTER FOUR

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 non-traditional variables, such as compensatory strategy use and environmental cognitive demands, can account for additional variance in everyday executive skills 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 everyday executive ability. Although several authors had speculated that these variables are important to ecological validity research, 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 DEX and similar measures of executive functioning (e.g., Burgess et al., 1998). Other studies, however, have reported null findings using the DEX and neuropsychological measures of executive functioning (Bogod et al., 2003; Chan, 2003; 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 varies 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.

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. In addition, the executive tests did not demonstrate significant discriminant validity, as the non-executive tests accounted for 21 percent of the variance and the executive tests failed to account for additional variance above and beyond this. The general neuropsychological measures also failed to account for additional variance above that accounted for by the executive tests. This indicates that the outcome measure may be assessing skills other than just executive functioning, or the general neuropsychological measures may involve more executive functioning than is assumed. In the correlational analysis, VIQ was highly related to both the DEX and the BAFQ. This suggests that either the measures of everyday executive functioning are partially measuring general intellectual ability, or VIQ is partially measuring executive functioning, or both.

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. Further, 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.

Because previous research has suggested that reliance on an overall "executive functioning" indicator may obscure more specific relationships between tests and certain aspects of the dysexecutive syndrome, we also analyzed the factor structure of the DEX and BAFQ. Factor analysis of the DEX revealed five factors. The first three factors, inhibition, goal-directed behavior/intentionality, and executive memory/cognition, were generally consistent with previous factor analytic studies of the DEX (Amieva et al., 2003; Burgess et al., 1998; Chan, 2001). The fourth factor (awareness) was consistent with the "knowing-doing dissociation" factor from the Chan study. The fifth factor (restlessness/hyperactivity) was not consistent with any previous factor solution. The current research provides additional support for at least three replicable factors from the DEX. Despite the fact that previous research on the factor structure of the BAFQ has found a two factor solution (Orbitofrontal and Dorsolateral), the current study revealed a four factor solution, with considerable overlap between the first three factors and the first three factors from the DEX. The previous research on the BAFQ has used normal or closed-head injury populations (Dywan et al., 1995; Simpson & Schmitter-Edgecombe, 2002), suggesting that the lack of correspondence between these factor solutions may be due to the different populations under study.

Providing further evidence that the factors generated from the DEX and BAFQ are indeed concordant, the first three factors from each outcome measure displayed a similar pattern of correlations with the neuropsychological tests. In particular, the executive memory/cognition factors from both the DEX and the BAFQ were more highly related to the neuropsychological tests than the other factors, and included correlations with neuropsychological tests of memory and verbal intelligence in addition to Trails B and the Stroop. A few exceptions were noted; however, as the inhibition factor from the BAFQ was related to the Stroop and the BNT, while the inhibition factor from the DEX was not. Thus, although these factors appear generally similar in terms of the items that load highly on them, they may be assessing slightly different abilities. Further evidence for this includes the fact that the fourth factor from the BAFQ and the fourth and fifth factors from the DEX appeared to measure different aspects of the dyexecutive syndrome.

Given that this study found evidence in support of separate executive factors, it may be more accurate to use these factor scores, rather than an overall score, when examining the ecological validity of executive tests. Consistent with findings using the overall score, the correlations between the factor scores from both the DEX and BAFQ and the neuropsychological measures revealed that only the Stroop and Trails B were significantly related to the factor scores. Thus, in this study, separating the outcome measure into factors did not improve the correlations of individual tests. In contrast, the group of executive tests did significantly predict the executive memory factor from the DEX, while they did not significantly predict the overall

DEX score. Further, when looking at each factor score from the BAFQ separately, the environmental cognitive demand improved prediction of all factors beyond the executive functioning tests, while compensatory strategy use improved prediction of only the intentionality and executive memory factors. Accounting for both variables improved the relationship between the executive tests and only the executive memory factor score. Thus, the executive memory component of the dysexecutive syndrome appears most highly related to formal neuropsychological testing. In addition, both compensatory strategy use and environmental cognitive demand appear to be related to the executive memory factor, while the latter variable was also significantly related to the other 3 factors. This suggests that environmental demand is important to all aspects of the dysexecutive syndrome, while executive functioning tests and compensatory strategy use may be more specific to certain aspects of the dysexecutive syndrome.

The results of the factor analyses as a whole suggest that the dysexecutive syndrome is fractionable, as previous research has indicated; however, the precise nature of these components requires additional study before they are used clinically in favor of an overall score. Further exploration of extra-test variables and their relationships to the various components of the dysexecutive syndrome is also important. Large scale factor analytic studies using several measures of everyday executive ability and neuropsychological measures are needed to provide a more reliable set of factors. Based on this and previous research, inhibition, intentionality and executive memory appear to be the most consistent distinct components of the dysexecutive syndrome as measured by the DEX and BAFQ. Further, the executive memory factor appears to be the most highly predicted by the group of executive functioning tests used in this study.

Limitations

There are several limitations to the current study. First, the sample size is relatively small. Therefore, the overall pattern of results should be interpreted, rather than any single finding. Where possible, only findings that were consistent for both the DEX and the BAFQ were emphasized in order to reduce the likelihood of spurious findings. Also, the number of tests examined was kept to a minimum. As a result, 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, these tests are commonly used in clinical practice and were thought to be of the most clinical relevance. Further, 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, selecting measures that included some non-executive components (i.e., Trails B and Stroop Color-Word) would actually reduce the likelihood of accounting for unique variance. If we had used more pure measures of executive functioning, it is likely that the executive tests would have accounted for even less variance. Variable and test selection is an important issue that has not been addressed well in the ecological validity literature.

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.

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 levels 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 = .58), and BAFQ (r = .61) 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 = -

.08). 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.

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, of the executive tests administered in this study, 2 out of 4 were not significantly related to either measure of everyday executive functioning. 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. Further, this set of tests was not uniquely related to the executive outcome measures, above and beyond that predicted by the non-executive tests. Second, the dysexecutive syndrome does not appear to be unitary, as at least three reliable factors have been found in this and other studies. More research is needed to further clarify if there are reliable additional factors. Lastly, the executive tests combined only 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.

REFERENCES

Amieva, H., Phillips, L. & Della Salla, S. (2003). Behavioral dysexecutive symptoms in normal aging. *Brain and Cognition, 53,* 129-132.

Bogod, N. M., Mateer, C. A. & MacDonald, S. W. (2003). Self-awareness after traumatic brain injury: A comparison of measures and their relationship to executive functions. *Journal of the International Neuropsychological Society,* 9, 450-458.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H. & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4,* 547-558.

Chan, R. (2001). Dysexecutive symptoms among a non-clinical sample: A study with the use of the Dysexecutive Questionnaire. *British Journal of Psychology, 92*, 551-565.

Chaytor, N. & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181-197.

Chelune, G. J. (1985). Toward a neuropsychological model of everyday functioning. *Psychotherapy in Private Practice, 3*(3), 39-44.

Clark, C., Prior, M. & Kinsella, G. (2000). Do executive function deficits differentiate between adolescents with ADHD and oppositional defiant/conduct disorder? A neuropsychological study using the six elements test and hayling sentence completion test. *Journal of Abnormal Child Psychology, 28*(5), 403-414.

Dimitrov, M., Grafman, J. & Hollnagel, C. (1996). The effects of frontal lobe damage on everyday problem solving. *Cortex, 32*, 357-366.

Dywan, J., Roden, R. & Murphy, T. (1995). Orbitofrontal symptoms are predicted by mild head injury among normal adolescents. *Journal of the International Neuropsychological Society, 1*, 121.

Dywan, J. & Segalowitz, S. (1996). Self and family ratings of adaptive behavior after traumatic brain injury: Psychometric scores and frontally generated ERPs. *Journal of head Trauma Rehabilitation, 11,* 79-95.

Evans, J. J., Chua, S. E., McKenna, P. J. & Wilson, B. A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine, 27*, 1-12.

Franzen, M. D. & Arnett, P. A. (1997). The validity of neuropsychological assessment procedures. In H. W. Reese & M. D. Franzen (Eds.), *Biological and neuropsychological mechanisms: Life-span developmental psychology*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Golden, C. (1978). *Stroop color and word test: a manual for clinical and experimental uses*. Los Angeles, CA: Western Psychological Services.

Goldstein, G. (1996). Functional considerations in neuropsychology. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing.* Delray Beach, FL: GR Press/St. Lucie Press.

Goldstein, G. & McCue, M. (1995). Differences between patient and informant functional outcome ratings in head-injured individuals. *International Journal of Rehabilitation and Health, 1*(1), 25-35.

Heinrichs, R. W. (1990). Current and emergent applications of neuropsychological assessment: Problems with validity and utility. *Professional Psychology: Research and Practice, 21*(3), 171-176.

Kaitaro, T., Koskinen, S. & Kaipio, M. (1995). Neuropsychological problems in everyday life: A 5-year follow-up study of young severely closed-head-injured patients. *Brain Injury, 9*(7), 713-727.

Kaplan, E., Goodglass, H. & Weintraub, S. (1983). *The Boston Naming Test.* 2nd ed., Philadelphia: Lea & Febiger.

Kortte, K., Horner, M. & Windham, W. (2002). The trail making test, part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology, 9*(2), 106-109.

Lezak, M. (1995). *Neuropsychological Assessment (3rd Ed).* New York: Oxford University Press.

Long, C. J. & Collins, L. F. (1997). Ecological validity and forensic neuropsychological assessment. In R. J. McCaffrey, A. D. Williams, J. M. Fisher & L. C. Laing (Eds.), *The practice of forensic neuropsychology: Meeting challenges in the courtroom*. New York: Plenum Press.

Long, C. J. & Kibby, M. Y. (1995). Ecological validity of neuropsychological tests: A look at neuropsychology's past and the impact that ecological issues may have on its future. *Advances in Medical Psychotherapy, 8*, 59-78.

McCue, M. & Pramuka, M. (1998). Functional assessment. In G. Goldstein & S. Beers (Eds.), *Rehabilitation.* New York: Plenum Press.

Norris, G. & Tate, R. L. (2000). The behavioural assessment of the dysexecutive syndrome (BADS): Ecological, concurrent and construct validity. *Neuropsychological Rehabilitation, 10*(1), 33-45.

Poole, J. H., Ober, B. A., Shenaut, G. K. & Vinogradov, S. (1999). Independent frontal-system deficits in schizophrenia: Cognitive, clinical, and adaptive implications. *Psychiatry Research, 85*, 161-176.

Psychological Corporation (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Harcourt Brace & Company.

Ready, R. E., Stierman, L. & Paulsen, J. S. (2001). Ecological validity of neuropsychological and personality measures of executive functions. *The Clinical Neuropsychologist, 15*(3), 314-323.

Sbordone, R. J. (1996). Ecological validity: some critical issues for the neuropsychologist. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 15-41). Delray Beach, FL: GR Press/St. Lucie Press.

Sbordone, R. J. (1997). The ecological validity of neuropsychological testing. In A. M. Horton, D. Wedding & J. Webster (Eds.), *The neuropsychology handbook, volume 1: Foundations and assessment (2nd Ed.).* New York: Springer Publishing Company.

Sbordone, R. J. & Guilmette, T. J. (1999). Ecological validity: Prediction of everyday and vocational functioning from neuropsychological test data. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice.* Lisse: Swets and Zeitlinger.

Simpson, A. & Schmitter-Edgecombe, M. (2002). Prediction of employment status following traumatic brain injury using a behavioural measure of frontal lobe functioning. *Brain Injury, 16*(12), 1075-1091.

Solanto, M. V., Abikoff, H., Sonuga-Barke, E., Schachar, R., Logan, G. D., Wigal, T., Hechtman, L., Hinshaw, S. & Turkel, E. (2001). The ecological validity of delay aversion and response inhibition as measures of impulsivity in AD/HD: A supplement to the NIMH multimodal treatment study of AD/HD. *Journal of Abnormal Child Psychology*, *29*(3), 215-228.

Spreen, O. & Strauss, E. (1998). *A Compendium of Neuropsychological Tests (2^{nd} Ed).* New York: Oxford University Press.

Sunderland, A., Harris, J. E. & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior, 22,* 341-357.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scales-III.* San Antonio, Texas: Psychological Corporation.

Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied and Preventive Psychology, 2,* 209-215.

Wilson, B. A., Alderman, N., Burgess, P., Emslie, H. & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome (BADS).* Bury St. Edmunds, England: Thames Valley Test Company.

Appendix A

Modified Dysexecutive Questionnaire

Participant's name___________________________ Rater's name____
Date of rating _______________________________ Relationship to participant _______________________________

This questionnaire looks at some of the difficulties that people sometimes experience. We would like you to read the following statements, and rate them on a five-point scale according to your experience of ___________________. Make sure you answer all 3 parts for each symptom. For Part B you should think about the everyday activities that he/she usually does and decide whether having each symptom would interfere with these activities or not. This is important because everyone's daily routine is different. For example, some people's daily routines require a lot of memory (e.g., student), while other people's usual activities require very little memory (e.g., factory worker).

For Part B: If you circled "0" in Part A, substitute the following question for Part B: "If he/she had problems in this area, how often would it interfere with his/her usual daily activities?"

The rating scale is as follows:

0 = Never
1 = Occasionally
2 = Sometimes
3 = Fairly Often
4 = Very Often

	Part A	Part B	Part C
Symptom:	How often does this problem occur?	**How often do (or would) problems in this area interfere with his/her usual daily activities?	How often does he/she do something to compensate for, or prevent, difficulties in this area?
1. Has problems understanding what other people mean unless they keep things simple and straightforward	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
2. Acts without thinking, doing the first thing that comes to mind	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
3. Sometimes talks about events or details that never actually happened, but he/she believes did happen	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
4. Has difficulty thinking ahead or planning for the future	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
5. Sometimes gets over-excited about things and can be a bit "over the top" at these times	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
6. Gets events mixed up with each other, and gets confused about the correct order of events	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
7. Has difficulty realizing the extent of his/her problems and is unrealistic about the future	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4

****For Part B: If you circled "0" in Part A, substitute the following question for Part B: "If he/she had problems in this area, how often would it interfere with his/her usual daily activities?"**

The rating scale is as follows:

0 = Never
1 = Occasionally
2 = Sometimes
3 = Fairly Often
4 = Very Often

Symptom:	Part A How often does he/she have this problem?	Part B **How often do (or would) problems in this area interfere with his/her usual daily activities?	Part C How often does he/she do something to compensate for this problem?
8. Seems lethargic, or unenthusiastic about things	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
9. Does or says embarrassing things when in the company of others	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
10. Really wants to do something one minute, but couldn't care less about it the next	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
11. Has difficulty showing emotion	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
12. Loses his/her temper at the slightest thing	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
13. Seems unconcerned about how he/she should behave in certain situations	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
14. Finds it hard to stop repeating saying or doing things once started	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
15. Tends to be very restless, and "can't sit still" for any length of time	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
16. Finds it difficult to stop doing something even if he/she knows he/she shouldn't	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
17. Will say one thing, but will do something different	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
18. Finds it difficult to keep his/her mind on something, and is easily distracted	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
19. Has trouble making decisions, or deciding what he/she wants to do	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
20. Is unaware of, or unconcerned about, how others feel about his/her behavior	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4

Appendix B

The Brock Adaptive Functioning Questionnaire

Name of Rater____________________________ Name of Patient________________________
Relationship to Patient___________________ Date of Rating_________________________

People are very different in the way they approach situations. Please answer each question based on the behavior of your friend or family member AT THIS TIME.

Circle the item that best describes his/her behavior. Read the choices carefully each time so you circle the right choice.

	0=hardly ever/never **1=rarely** **2=sometimes** **3=often** **4=almost always**

Planning:

Does he/she have a hard time making plans for the day on their own?	0 1 2 3 4
When going out for the day, does he/she think about what might be needed later in the day, for example, bringing a jacket in case it got colder?	0 1 2 3 4
When he/she has several tasks to do, does he/she organize them in an efficient way?	0 1 2 3 4
Would he/she be able to manage (take appropriate steps) if an emergency came up and they were home alone?	0 1 2 3 4
When making choices, does he/she consider how these choices may affect them in the future?	0 1 2 3 4
When making long-term plans, does he/she consider how these choices may affect them in the future?	0 1 2 3 4
When he/she makes plans, would you say that their plans show good judgment (i.e., are they workable and realistic?)	0 1 2 3 4

Initiation:

Is he/she able to get up on time in the morning without actually being prompted by another person?	0 1 2 3 4
Does he/she carry out their household jobs without being reminded by anyone?	0 1 2 3 4
Does he/she have trouble getting started on a project unless someone starts them off?	0 1 2 3 4
Even though he/she knows exactly what has to be done to keep a project going, does he/she have a hard time moving to the next step on their own?	0 1 2 3 4

	0=hardly ever/never **1=rarely** **2=sometimes** **3=often** **4=almost always**

Flexibility:

Once he/she has made plans, do they find it very difficult to change them?	0 1 2 3 4
When doing a task, can he/she easily distinguish between the more important and the less important aspects of the task. (That is, if forced to hurry, would he/she be able to skip the less important steps?)	0 1 2 3 4
When telling someone about an event or a movie, can he/she easily skip unimportant details if pressed for time?	0 1 2 3 4
Does he/she have a hard time switching topics during conversation?	0 1 2 3 4

Excess Caution:

Does he/she appear to go over and over the same things in their mind more than they need to?	0 1 2 3 4
Does he/she like to do things in the same way each time?	0 1 2 3 4
Does he/she become uncomfortable if their usual routines have to be changed?	0 1 2 3 4
Does he/she check many times to make sure that things are safe (e.g., door locked, stove off etc.)?	0 1 2 3 4
Does he/she seem more suspicious of other people than you think is necessary?	0 1 2 3 4

Attention:

Does he/she get distracted easily?	0 1 2 3 4
Is he/she likely to forget that the stove or kettle has been left on?	0 1 2 3 4
Does he/she have a lot of trouble keeping track of where things are around the house?	0 1 2 3 4
Does he/she have trouble following spoken directions?	0 1 2 3 4
Does he/she have trouble sticking to the point that they are trying to make when having a discussion?	0 1 2 3 4
Is he/she easily confused in stores and shopping malls?	0 1 2 3 4
Is he/she likely to get lost even in relatively familiar places?	0 1 2 3 4

	0=hardly ever/never **1=rarely** **2=sometimes** **3=often** **4=almost always**

Memory:

Does he/she have a hard time learning new skills?	0 1 2 3 4
Does he/she have difficulty remembering events that happened in the last week?	0 1 2 3 4
Does he/she have difficulty remembering to do things they had planned to do?	0 1 2 3 4
Does he/she have trouble remembering the names of people that they see regularly?	0 1 2 3 4
Does he/she have a hard time recognizing people they have met before?	0 1 2 3 4
Does he/she have trouble recalling things that they used to know quite well?	0 1 2 3 4
Does he/she tell people things that may not be true?	0 1 2 3 4
Does he/she tell people things that could not possibly be true?	0 1 2 3 4

Arousal Level:

Does he/she have difficulty staying awake or alert?	0 1 2 3 4
Does his/her voice sound flat compared to other people?	0 1 2 3 4
Does he/she find it very difficult to get enthusiastic about things?	0 1 2 3 4
Does he/she find it very hard to maintain interest in what they are doing for a long period of time?	0 1 2 3 4
Does he/she seem very sad or depressed?	0 1 2 3 4

	0=hardly ever/never **1=rarely** **2=sometimes** **3=often** **4=almost always**

Emotionality:

Does he/she get much too excited about things?	0 1 2 3 4
Does he/she have difficulty controlling emotional responses e.g., crying much too easily?	0 1 2 3 4
Are there times when he/she laughs or talks too much or too loudly compared to others?	0 1 2 3 4
Do you find that his/her eye contact can be too intense during conversation?	0 1 2 3 4

Impulsivity:

Does he/she make inappropriate comments or blurt things out that would be better left unsaid?	0 1 2 3 4
Does he/she use alcohol (or other drugs) more than they should?	0 1 2 3 4
Does he/she spend money unnecessarily without giving it much thought?	0 1 2 3 4
Does he/she make sexual remarks which seem inappropriate?	0 1 2 3 4
Does he/she touch people in ways which are sexually inappropriate?	0 1 2 3 4
Does he/she have a lot of trouble controlling the amount they eat?	0 1 2 3 4
Does he/she need external constraints in order to control eating (for example, careful control over what food is left around the house, or a lock on the refrigerator)?	0 1 2 3 4

Aggressiveness:

Is he/she quick to take offense at what others say?	0 1 2 3 4
When he/she gets frustrated, will they throw things around or damage things?	0 1 2 3 4
When he/she gets angry, will they threaten people?	0 1 2 3 4
If pushed to the limit, could he/she strike out at someone?	0 1 2 3 4
Would he/she do what they really want to do, even if it is illegal?	0 1 2 3 4

	0=hardly ever/never **1=rarely** **2=sometimes** **3=often** **4=almost always**

Social Monitoring:

Does he/she stand a little too close when engaged in a conversation?	0 1 2 3 4
Does he/she seem to miss the point of many jokes or stories that other people seem to enjoy?	0 1 2 3 4
Does he/she seem to pay attention to whether others are following what they are saying?	0 1 2 3 4
When he/she is telling things to other people, does he/she give them as much background information as is needed so people can follow easily?	0 1 2 3 4
If others are looking disinterested in what he/she is saying, will he/she try to stop talking or change the topic?	0 1 2 3 4
Does he/she tend to tell the same story over again to the same people?	0 1 2 3 4
When a social situation has gone badly, does he/she try to figure out what went wrong so they can make it go better the next time?	0 1 2 3 4

Empathy:

If he/she just received a favor or some special consideration, would they show appropriate appreciation?	0 1 2 3 4
Would he/she notice if someone was feeling overtired or worried?	0 1 2 3 4
If he/she noticed that someone close looked overworked or worried, would they do what they could to ease their load?	0 1 2 3 4
Does he/she seem to notice when other people are feeling awkward in a social situation?	0 1 2 3 4
When someone is feeling awkward in a social situation, will he/she do something that makes the person feel more comfortable?	0 1 2 3 4

**Are there any other areas in everyday functioning where your family member or friend does very well or very poorly that have not been covered in this questionnaire that you think would be important to mention? Please describe below: