VALIDITY AND DIAGNOSTIC ACCURACY OF A CLINICAL TRIAL BATTERY FOR PRIMARY BRAIN TUMOR PATIENTS

_____

A Dissertation

Presented to

The Faculty of the Department

of Psychology

University of Houston

_____

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

_____

By

Surabhi Y. Patwardhan

December, 2012

VALIDITY AND DIAGNOSTIC ACCURACY OF A CLINICAL TRIAL BATTERY FOR PRIMARY BRAIN TUMOR PATIENTS

_____

An Abstract of a Dissertation

Presented to

The Faculty of the Department

of Psychology

University of Houston

_____

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

_____

By

Surabhi Y. Patwardhan

December, 2012

ABSTRACT

Objective: Neurocognitive function (NCF) is a relevant endpoint in clinical trials of primary brain tumor patients. The Clinical Trial Battery (CTB) consisting of the Hopkins Verbal Learning Test-Revised, Trail Making Test, and Controlled Oral Word Association is currently in use for multinational / multisite clinical trials of primary brain tumor patients. Although the CTB is brief, useful, and consists of standardized measures, validity and diagnostic accuracy of the CTB are unknown and the measures are primarily verbally based. The CTB generates multiple outcome variables for cognitive domains of attention, executive function, learning, memory, and processing speed, and can therefore be analytically challenging. The present study aimed to investigate the validity and diagnostic accuracy of the CTB and to evaluate the added value of including a test requiring predominantly visuospatial function.

Methods: To integrate and interpret the six major outcome variables generated by the CTB, an unweighted average standard composite score (NCF6Z) was calculated. Block Design (BD) was added to the CTB in order to examine the impact of including a test of visuospatial function, which yielded another composite score (NCF7Z) for the extended CTB (CTBE). Validity and diagnostic accuracy of NCF6Z and NCF7Z of untreated primary brain tumor patients (n = 260) were assessed against the criteria: performance on comprehensive clinical assessment, clinician ratings of NCF impairment status, tumor characteristics, and clinician report measures of functional status.

Results: NCF6Z was found to be adequately valid. Assessed against clinician ratings of NCF impairment status, at cutpoint -0.50, NCF6Z demonstrated sensitivity = 0.86, specificity = 0.78, PPV = 0.80, and NPV = 0.85. At cutpoint -0.6257, NCF6Z displayed optimal

diagnostic accuracy (sensitivity = 0.85, specificity = 0.83, PPV = 0.84, NPV = 0.84). Addition of BD did not augment validity or diagnostic accuracy of the original CTB, even when the scores were weighted to offset the purported verbal bias of the CTB.

Conclusion: The present investigation demonstrated validity and diagnostic accuracy related evidence for NCF6Z. The CTB, which is a brief, useful, valid, and diagnostically accurate measure of NCF among primary brain tumor patients, is suitable for use in large scale research studies involving primary brain tumor patients.

ACKNOWLEDGEMENTS

I sincerely thank patients at The University of Texas M. D. Anderson Cancer Center, Houston, TX, whose data were used in this study and made this work possible. I am grateful to Dr. Jeffrey S. Wefel and Dr. Paul J. Massman for their continuous guidance, advice, and support in my educational progress and achievement of professional goals.

I would also like to thank my dissertation committee members: Dr. Merrill Hiscock and Dr. Karla K. Stuebing. I am thankful to Dr. Anne Herring, Dr. Nicholas J. K.Breitborde, and Dr. Ron Wright from the Department of Psychiatry, The University of Arizona Medical Center, Tucson, AZ for their guidance and encouragement. I would like to put on record help extended by the administrative staff: Dr. Suzanne Kieffer, Ms. Patti Tolar, Ms. Caroline Watkins, and Mr. Joshua Watkins at the Department of Psychology, University of Houston, Houston, TX. I thank my colleagues and friends Dr. Eric Waldron, Dr. Shruti Rane, Dr. Mekhala Samsi, Dr. Kara Fitzgerald, and Dr. Irene Teo, who were there. I am grateful to all my teachers who gave me a solid foundation and led me to this day.

Finally, I thank my husband, Dr. Viraj Pandit, my parents Mr. Yeshwant and Mrs. Amita Patwardhan, my sister Dr. Tanvi Patwardhan, my uncle Mr. Anant Danait, and the rest of my family for their unwavering support, and non-judgmental love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Validity and Diagnostic Accuracy of a Clinical Trial Battery for Primary Brain Tumor Patients

Neurocognitive dysfunction is a common adverse symptom among brain tumor patients and may reflect tumor burden and/or treatment effects (Wefel, Kayl, & Meyers, 2004). Neurocognitive function (NCF) has been established as a predictor of patient quality of life (Buckner, O'Fallon, Iturria, et al., 2004), tumor progression (Klein, Postma, Taphoorn, et al., 2003, Meyers & Hess, 2003, Meyers, et al., 2000) and patient survival (Johnson, Sawyer, Meyers, et al., 2012, Buckner et al., 2004; Meyers Hess, Yung, et al., 2000; Taphoorn et al., 2004). When present, neurocognitive dysfunction influences patient functional status, which includes successful engagement in activities of daily living (Meyers, Hess, Yung, et al., 2000; Meyers, Smith, Bezjak, et al., 2004; Taphoorn & Klein, 2004). Improving, stabilizing and/or slowing the decline in NCF is, thus, an important goal in clinical care as well as a highly relevant endpoint in clinical trials (Meyers & Brown, 2006) of brain tumor patients. In addition, NCF is a useful outcome to assist in measuring the net clinical benefit of cancer therapy, which includes "beneficial effects on disease related symptoms and/or quality of life" (Meyers et al., 2006).

The Clinical Trial Battery:

Meyers et al. (2006) recommend that a clinical trial battery consist of brief, psychometrically sound, objective and standardized tests with published normative data that are simple to administer by appropriately trained staff and have been shown to be sensitive to

the neurotoxic effects of cancer and cancer therapies. At present, The Clinical Trial Battery (CTB) is the neuropsychological assessment battery in use for numerous multisite / multinational oncology clinical trials assessing neurocognitive endpoints. The CTB consists of the following NCF tests: Trail Making Test Part A and Part B (TMTA and TMTB), Hopkins Verbal Learning Test–Revised (HVLT-R)—Total Recall (HVLT-R TR), Delayed Recall (HVLT-R DR), Delayed Recognition (HVLT-R DRECOG), and Controlled Oral Word Association (COWA). The CTB takes 25–30 minutes for complete administration, which is critical for the oncology clinical trial environment, because it allows for reduced clinician burden, is cost and time efficient, and decreases training demands on the examiner (Carey, Woods, Rippeth, et al., 2004a; Meyers et al., 2006). Consistent with the recommendations of Caine, Mehta, Laack, et al. (2012), the CTB also requires minimal special materials or apparatus.

The CTB measures cognitive domains that are commonly affected by tumor and treatment including attention, executive function, learning and memory, and verbal fluency (Fox, Michael, & Booth-Jones, 2006, Wefel et al., 2004, Klein et al., 2003, Weitzner & Meyers, 1997). In contrast, some previous attempts at developing neurocognitive assessment batteries for clinical trials were criticized for their narrow scope (Weitzner et al., 1997). For example, Grant, Slattery, Gregor, Cull, Traynor, et al. (1994) proposed a brief, repeatable battery for glioma patients, consisting of measures of motor function (timed nine-hole peg test, 10 meter walk), short term memory (Williams delayed recall test), and language (Boston Aphasia Severity Rating Scale). This battery failed to assess cognitive domains important for functional independence, such as speed of processing and executive function (Weitzner et al., 1997).

Psychometric and Domain Related Limitations of the CTB:

While the CTB consists of clinically relevant, psychometrically sound, standardized, and a simple-to-administer set of individual tests, it suffers from some limitations. There is no information on psychometric characteristics of the CTB as an assessment battery. For example, no validity related evidence on the CTB is currently available. Secondly, the CTB consists of a group of standardized and psychometrically sound measures that generate six outcome variables: TMTA, TMTB, COWA, HVLT-R TR, HVLT-R DR, and HVLT-R DRECOG. Evaluation and interpretation of multiple outcomes within a single battery is analytically challenging. Risk of error increases with an increasing number of outcome measures (Ingraham & Aikken, 1996). Similarly, Weitzner et al. (1997) noted that a neurocognitive assessment battery should be able to classify patients based on their neurocognitive status. This ability to correctly classify patients into clinically relevant groups reflects diagnostic accuracy (Zweig & Campbell, 1993). Sensitivity, specificity, positive predictive value, and negative predictive value are indices of diagnostic accuracy of an instrument. Data on diagnostic accuracy of the CTB are lacking.

Concerns have been raised that the CTB is predominantly verbally based. It is unclear if a nonverbal visual constructional measure would enhance the CTB. In some studies, visuospatial function, along with verbal memory and psychomotor speed, has been found to be sensitive to neurocognitive impairments among primary brain tumor patients (Lageman, Cerhan, Locke, et al., 2010).

Overcoming the Psychometric and Domain Related Limitations of the CTB:

In order to examine the impact of including a non-verbal / visuospatial test on the CTB, we added Block Design from the Wechsler Adult Intelligence Scale—Third edition (BD, WAIS-III) to the CTB, which we then renamed as the CTB-extended (CTBE). Similar to the subtests of the CTB, BD is a frequently used measure in neuropsychological evaluations of Primary Brain Tumor patients.

Accurate representation of the multiple outcome variables of the CTB is required in order to assess validity and diagnostic accuracy of the CTB. As noted, multiple outcome variables are analytically challenging and may best be represented through a single summary score. Two types of summary scores have been used in the previous studies, namely Global Deficit Scores (GDS) and Composite Scores.

Global Deficits Scores:

GDS account for the number and severity of deficits in an individual's performance throughout the test battery (Carey, Woods, Gonzalez et al., 2004; Miller & Rohling, 2001). The GDS, created by Heaton, Grant, and Matthews (1991), is based on conversion of demographically corrected T-scores on individual neuropsychological measures to deficit scores (see Table 1) ranging from 0 (no impairment) to 5 (severe impairment) (Carey, Woods, Gonzalez, et al., 2004b). The deficit scores are then averaged to create the GDS that weights deficient performances higher than those within or above the normal limits (Heaton et al., 1994a; Heaton, Paulsen, McAdams, et al., 1994b; Heaton et al., 1995).

Table 1. Conversion of T-scores into Deficit Scores

T scores	Deficit Score	Impairment Descriptor
≥40	0	Normal
39-35	1	Mild
34-30	2	Mild-to-Moderate
29-25	3	Moderate
24-20	4	Moderate-to-severe
≤ 19	5	Severe

Carey et al. (2004a) assessed 88 Human Immunodeficiency Virus seropositive (HIV+) patients and 61 HIV- participants with a battery consisting of the following component measures: Hopkins Verbal Learning Test–Revised, Brief Visuospatial Memory Test – Revised, FAS verbal fluency, Animal Fluency, Stroop Color-Word Test, Trail Making Test, Wisconsin Card Sorting Test – 64 Card Version, Halstead Category Test, Paced Auditory Serial Addition Test, Grooved Pegboard, Digit Symbol, Symbol Search, and Letter-Number Sequencing tests (WAIS-III). The GDS was able to discriminate between HIV+ and HIV- participants. For cutoff values of ≥ 0.25 to ≥ 2.00, specificity of the GDS ranged between 0.71 to 1.00 and positive predictive power ranged between 0.74 to 1.00. There was a significant correlation between the GDS and blind clinician ratings of global NCF impairment ($\rho = 0.87$, $p < .0001$). Against the standard of clinical ratings, the GDS was able to identify HIV+ individuals with NCF impairments (specificity range = 0.69 to 1.00 and positive predictive power = 0.71 to 1.00). Other studies have established that the GDS is useful in detecting mild cognitive impairment in HIV+ patients and is related to the biological markers of HIV-associated immunosuppression (Gonzalez, Heaton, Moore et al., 2003; Heaton, et al., 1994a).

Composite scores:

Another statistical approach to integrating and interpreting data from multiple measures of a neurocognitive battery is the development of a summary composite score based on individual scores of component measures. The composite score is usually the average of the standardized scores of all individual measures of the neurocognitive battery.

In order to assess the diagnostic accuracy of a neurocognitive assessment battery in detecting HIV associated dementia among HIV+ patients, Bottiggi, Chand, Schmitt, et al. (2007) created a composite score termed NPZ8. The NPZ8 was the average of eight z-scores obtained from each of the eight component tests of the neurocognitive battery: Symbol Digit Modalities, Grooved Pegboard (dominant and non-dominant), Trail Making Test Part A and B, Rey Auditory Verbal Learning Test Trial 8, Rey Auditory Verbal Learning Test Total, and COWA. The criterion for impairment ("cognitively impaired / demented") was established as an NPZ8 score of ≤ -2.00. Another criterion of impairment was based on individual test scores such that patients obtaining z-scores of ≤ -2.00 SD from the normative mean on two or more individual neuropsychological tests (e.g., Grooved Pegboard and COWA) were also classified as "cognitively impaired." If patients obtained one score of ≤ -2.00 SD on a single individual neuropsychological measure, then they were considered cognitively normal / subclinical. When compared to clinical ratings on the Memorial Sloan Kettering (MSK) staging scale, the NPZ8 exhibited a sensitivity of 0.30 and a negative predictive value of 0.58.

The Multiple Sclerosis Functional Composite (MSFC) (Cutter, Baier, & Rudick, 1999) is another composite score that is a focused and sensitive measure of disability among patients with multiple sclerosis (MS). It is a quantitative instrument that consists of three

measures: arm/hand dexterity (the 9-Hole Peg Test), leg function (the Timed 25-Foot Walk), and cognition (the Paced Auditory Serial Addition Test, 3-second version). The mean of the scores of the 2 Timed 25-Foot Walk trials, the reciprocal of the average score of the 2 trials of each hand, and the number of correct responses on the Paced Auditory Serial Addition Test were used to derive the MSFC score, which is the mean of the z scores of the three component scores. Cohen, Cutter, Fischer, et al. (2001) reported that the MSFC is simple to administer, has a practical testing protocol, and exhibits sound intra-rater and inter-rater reliability. Miller, Rudick, Cutter, et al. (2000) reported that the MSFC correlated with disability as measured by the Kurtzke Expanded Disability Status Scale (EDSS) ($r = -0.0.80$, $p < .001$), patient report of physical functioning (SIP Physical Summary Scale: $r = -0.71$, $p < .001$, SF36 Physical Component Score: $r = -0.41$, $p < .001$), and emotional functioning (SIP Psychosocial Summary Scale: $r = -0.34$, $p < .001$). Kalkers, Bergers, de Groot, et al. (2001) showed that the MSFC correlated significantly with the biologic disease markers of MS: T1-hypointense lesion load ($r = -0.24$, $p < .01$) and T2-weighted lesion load ($r = -0.25$, $p < .01$).

Composite scores have also been used in schizophrenia research (For example, Tabarés-Seisdedos, Balanzá-Martínez, Sánchez-Moreno, et al., 2008, Keefe, Bilder, Harvey et al., 2006). Keefe, Goldberg, Harvey et al. (2004) calculated an NCF composite score for the Brief Assessment of Cognition in Schizophrenia (BACS) by averaging the six standardized primary measures generated by the BACS, and then obtaining a z-score of the composite score. Test-retest reliability coefficients ranged between 0.86 to 0.95 for both patients with schizophrenia as well as healthy controls. Among patients with schizophrenia, the BACS measures correlated at least moderately with all NCF domains assessed by a comprehensive standard battery.

Heaton et al. (1994a) criticized composite scores for being arbitrary and of questionable sensitivity. They observed that composite scores have been used in various studies assessing neurocognitive impairment among HIV seropositive patient population. In these studies, HIV+ patients were classified as "impaired", if the composite scores were one or two standard deviations below the means of reference groups. However, there was no prior validation of the cut-off criterion. Moreover, in many cases, demographic and clinical characteristics of the reference groups differed from the research population. As a result, Heaton et al. (1994a) questioned the sensitivity as well as clinical relevance of the classifications based on composite scores. Nevertheless, composite scores are convenient, easy-to-calculate, and avoid multiple comparisons and thus correct for interdependence of NCF measures (Réthelyi, Czobor, Polgár, et al., 2012). Composite scores have been shown to be adequately valid, diagnostically accurate, and practically useful in previous studies (for example, Bottiggi et al, 2007, Keefe et al., 2004, Miller et al., 2001, Kalkers et al., 2001). As a result, composite scores are a commonly used indicator of global NCF assessed by a neurocognitive assessment battery.

Development of NCF6Z:

Among primary brain tumor patients with heterogeneous lesion locations, a composite score was expected to allow deficits on different measures to contribute equally to the impairment index. Further, in clinical trials it is less complex to analyze a single outcome measure. Based on the composite score methods previously reviewed that derived the average z-score from a battery of tests, we calculated the "NCF6Z" score. NCF6Z is the average of the individual z-scores of the 6 outcome variables (TMTA, TMTB, COWA,

HVLTR-TR, HVLTR-DR, and HVLTR-DRECOG) of the CTB. Below is the formula used to calculate NCF6Z:

$$NCF6Z = \frac{HVLTRTR + HVLTRDR + HVLTDRECOG + COWA + TMTA + TMTB}{6}$$

There was an additional outcome variable (BD) for the CTBE. Thus, the CTBE generated a composite score called NCF7Z. Below is the formula used to calculate NCF7Z:

$$NCF7Z = \frac{HVLTRTR + HVLTRDR + HVLTDRECOG + COWA + TMTA + TMTB + BD}{7}$$

Validation Criteria for NCF6Z:

The standard core comprehensive neuropsychological assessment battery used at The University of Texas MD Anderson Cancer Center (UTMDACC) for clinical care of brain tumor patients is represented in Table 2. The CTB and the CTBE were derived from this battery.

Table 2. List of Measures in the CTB and Comprehensive Battery

Domain	Comprehensive Battery	The CTB	Abbreviation
Attention	WAIS-III Digit Span		DS
Processing Speed	WAIS-III Digit Symbol		DSym
	Trail Making Test—Part A	Trail Making Test—Part A	TMTA
Executive Function	Trail Making Test—Part B	Trail Making Test—Part B	TMTB
	Controlled Oral Word Association	Controlled Oral Word Association	COWA
	WAIS-III Similarities		SIM
Learning	Hopkins Verbal Learning Test—Revised Total Recall	Hopkins Verbal Learning Test—Revised Total Recall	HVLT-R TR
Memory	Hopkins Verbal Learning Test—Revised	Hopkins Verbal Learning Test—Revised	
	Delayed Recall	Delayed Recall	HVLT-R DR
	Delayed Recognition	Delayed Recognition	HVLT-R DRECOG
Visuospatial Function	WAIS-III Block Design		BD
Motor Dexterity	Lafayette Grooved Pegboard		
	Dominant		Peg-D
	Non-Dominant		Peg-ND
Language	Boston Naming Test		BNT
	Visual Naming Test		VNT
	Token Test		Token Test
Gross Motor Function	Grip Strength		
	Dominant		Grip-D
	Non-Dominant		Grip-ND

Based on the wider scope and greater number of measures of the comprehensive battery, it is expected that the comprehensive battery provides a relatively more accurate impairment status of brain tumor patients. To the extent that the CTB coincides with the comprehensive battery, the CTB is likely to provide adequate data on the neurocognitive condition of the patients. This would establish criterion related validity of the CTB. Thus, we obtained a composite score (NCFTOTALZ) for the standard core comprehensive battery

to compare the CTB against the criterion of the standard core comprehensive battery. We hypothesized that (H1): NCF6Z and NCF7Z would correlate significantly with NCFTOTALZ.

Additional evidence for criterion related validity of the CTB can be obtained by examining the relationship between NCF6Z and tumor characteristics of grade, laterality, and caudality. In this regard, based on previous studies (for example, Miotto, Junior, Silva, et al., 2011) that have found significant differences in NCF performances of low grade and high grade glioma patients, we hypothesized that (H2): patients with high grade tumors (Glioblastoma Multiforme, Anaplastic Astrocytoma, and Anaplastic Oligodendroglioma) would obtain lower NCF6Z scores when compared to those with low grade tumors. Given that the CTB is heavily verbally based, we hypothesized that (H3): patients with left hemisphere tumors would obtain significantly lower NCF6Z scores compared to patients with right hemisphere tumors. However, since the addition of BD would likely identify more cognitive deficits in patients with right hemisphere tumor, (H4): we hypothesized that NCF7Z will not be significantly different based on tumor laterality. Additionally, we hypothesized that (H5): patients with anterior tumors would obtain significantly lower NCF6Z compared with those with posterior tumors, while (H6) NCF7Z would not significantly differ between patients with anterior and posterior tumors.

We expected that NCF6Z would be sensitive to comorbid conditions that may be the result of tumor, treatment, or both. Given that seizures are a common neurologic comorbidity that affects one third of the patients with primary brain tumors (Rajneesh & Binder, 2009) and, in fact, are the first symptom of brain tumors in 30 to 90% of patients (Rasmussen, 1957), we investigated the relationship between tumor-related epilepsy and

NCF6Z. Previous evidence has shown that among low grade glioma patients, tumor-related epilepsy burden is associated with reduced NCF functioning in all NCF domains (Klein, Engelberts, van de Ploeg et al., 2003). Thus, we hypothesized that patients with history of tumor-related epilepsy would obtain lower NCF6Z and NCF7Z scores compared to patients without tumor-related epilepsy (H7).

Concurrent validity related evidence can be obtained by investigating relationships between NCF6Z and measures of functional status and affective distress. The Karnofsky Performance Scale (KPS) and Functional Independence Measure (FIM, 1987) are clinician report measures of functional status that have been commonly used with primary brain tumor patients. Several studies assessing functional outcomes of interventional strategies with primary brain tumors have utilized the KPS and the FIM to evaluate clinician-reported functional changes (for example, Marciniak, Sliwa, Spill, et al., 1996, Weitzner, Meyers, & Byrne, 1996). We hypothesized that NCF6Z and NCF7Z would significantly correlate with the KPS score (H8), and that NCF6Z and NCF7Z would significantly correlate with the FIM score (H9).

Lastly, primary brain tumor patients report depression and anxiety in reaction to the disease (Litofsky, Farace, Anderson et al., 2004, Anderson, Taylor, & Whittle, 1999). Specifically, Taphoorn et al. (2004) observed that high grade glioma patients report higher levels of depression, panic, anxiety, and fear of dying compared to low grade glioma patients. Commonly observed mood disturbances in primary brain tumor patients may impact attention and motivation (Anderson et al., 1999), which may subsequently reduce performance on several other cognitive domains. Given its complex nature, we explored the relationship of NCF6Z and NCF7Z with affective distress among primary brain tumor

patients. It was expected that NCF6Z and NCF7Z would correlate negatively with self-report of affective distress.

Assessing the Diagnostic Accuracy of the CTB:

Assessing the diagnostic accuracy of a measure requires a criterion against which the decision regarding the presence or absence of the condition of interest is made. In many cases, a previously established test is used as the diagnostic standard (for example, Miller et al., 2001), along with base rates of the condition of interest. In some other cases, alternative approaches, such as clinician ratings (for example, Carey et a., 2004a), have been utilized.

Problems with Base Rates as the Diagnostic Standard:

Base rates of NCF impairment among primary brain tumor patients are currently unknown. Several clinical, methodological, and practical difficulties present themselves in obtaining accurate base rates of NCF impairment in primary brain tumors. Among untreated primary brain tumors, cognitive deficits may be associated with focal signs (such as aphasia) due to tumor invasion, generalized symptoms (sixth nerve palsy, headaches) related to cerebral edema and increased intracranial pressure (DeAngelis, 2001), comorbid condition(s) (tumor-related seizure disorder), treatment of the comorbid condition(s) (antiepileptic medicines), emotional distress associated with a cancer diagnosis, or a combination of any of the above factors (Taphoorn et al., 2004, Wefel et al., 2004). Moreover, NCF impairment among primary brain tumor patients is further influenced by such factors as rate of tumor progression (Bosma, Vos, Heimans, et al. 2007, Meyers et al., 2000).

Rudimentary estimates suggest that 10% patients with low grade glioma and 40% to 60% patients with high grade gliomas present with abnormalities of mental status (DeAngelis, 2001). Heimans & Reijneveld (2012) reported that among high grade glioma patients, severe NCF impairments have been observed in as many as 89% patients. Many patients with low grade glioma also present with NCF declines (Heimans et al., 2012). Several studies have shown that attention, processing speed, memory, and executive function are some of the domains that are commonly impacted by brain tumors (For example, Wefel et al., 2004, Klein et al., 2003). However, accurate base rates of global NCF as well as domain specific impairment are lacking. In addition, no NCF measure or battery has been uniformly and consistently used in research studies of NCF among primary brain tumor patients and various studies have used different NCF instruments (Fox et al., 2006, Weitzner et al., 1997). As a result, for the purposes of the present study, clinician ratings were used as the diagnostic standard against which the accuracy of NCF6Z was assessed.

Clinician Ratings of Impairment:

Clinician ratings of global NCF status consider the full range of test scores generated by the NCF battery. Thus, clinician ratings are often useful in detecting impairment based on multiple measures of an NCF battery. Ratings are usually obtained from two or more clinicians extensively trained in using the clinical ratings system. Detailed, explicit, manualized, and standardized guidelines for assigning ratings decreases subjectivity in judgment, thus increasing the inter-rater reliability of ratings (Garb & Schramke, 1996). Recruiting clinicians with diverse level of experiences increases the external validity of the ratings (Garb et al., 1996).

Heaton, Grant, Butters, et al. (1995) obtained blind clinician ratings of the neurocognitive status of HIV+ patients using a nine-point scale ranging from one (above average) to nine (severely impaired). Ratings were assigned for each of seven neurobehavioral domains as well as for global neurocognitive status. A cut-off score of five was used to indicate definite mild cognitive impairment (Heaton et al., 1995). If patients exhibited impairment in two or more domains, they were classified as neurocognitively impaired. Heaton, Kirson, Velin, et al. (1994a) reported excellent inter-rater reliability of the clinical rating system ($\kappa = 0.84$). In addition, Carey et al. (2004a) reported a multitude of studies that have established the reliability and validity of clinical ratings both for diagnostic purposes and for detecting milder neurocognitive impairment in various neurological and medical conditions (for example, Filley, Heaton, Thompson, et al., 1990; Grant, Adams, & Reed, 1979; Grant et al., 1978, 1987a, 1987b; Grant, Heaton, McSweeney, Adams, et al., 1982; Heaton, Grant, Anthony, et al., 1981; Heaton, Grant, & Matthews, 1991; Heaton, McSweeney, Grant, Adams, et al., 1983; Heaton, Nelson, Thomson, et al., 1985; Heaton et al., 1994a, 1995; Reitan, 1974; Ryan, Adams, Heaton, et al., 1991). Thus, although Carey et al. (2004) report that clinician ratings have been criticized for being subjective and logistically difficult, clinician ratings are commonly used to detect neurocognitive impairment. When compared with other methods of data reduction that involve group means, clinician ratings have been found to be superior and of higher sensitivity.

In the present investigation, assessed against the clinician ratings, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated and utilized as indicators of the diagnostic accuracy of NCF6Z and NCF7Z. Sensitivity is the proportion of cases that a test identifies as positive, when in fact, the

condition of interest is present (Zweig et al., 1993). In contrast, specificity refers to the proportion of cases that a test identifies as negative, when in fact, the condition of interest is absent. PPV is the proportion of cases in which the condition of interest is present, when the test result is positive, while NPV is the proportion of cases in which the condition of interest is absent, when the test results are also negative (Zweig et al., 1993). We hypothesized that the CTB will have adequate sensitivity, specificity and further investigated the PPV, and NPV (H10). We also hypothesized that the CTBE will have adequate sensitivity, specificity, and further investigated the PPV, and NPV (H11). When the CTB would be utilized as a screening instrument, where identifying patients with NCF impairment is of chief importance, high sensitivity and NPV would be more desirable over specificity and PPV. When used as an outcome measure, high PPV may be preferred.

We assessed the sensitivity, specificity, PPV, and NPV of the CTB and CTBE across composite scores $\leq$ -0.50, -0.60, -0.75, -1.00, -1.25, and -1.50. Composite score of -0.50 was studied because it is roughly equivalent to an average score of -1.0 on approximately one-half of the component measures (Carey et al., 2004a). Similarly, z-scores $\leq$ -1.50 on a neurocognitive test are traditionally used to identify impaired ability in clinical settings.

We further generated Receiver Operating Characteristic (ROC) curves for NCF6Z, NCF7Z, and NCFTOTALZ (McNeil & Hanley, 1984). The ROC curves afforded us the ability to assess diagnostic accuracies of the composite scores across a range of cutpoints. Furthermore, using the Youden Index (Youden, 1950), also known as the *J* statistic, optimal cutpoints that provided ideal sensitivity, specificity, PPV, and NPV were identified. We then compared the ROC curves of the composite scores using area under the curve (AUC) (Hanley & McNeil, 1983).

Method

Participants:

The Section of Neuropsychology at The UTMDACC (Houston, TX) has followed a cohort of primary brain tumor patients longitudinally with detailed clinical and neuropsychological data at various times across their disease course. These data have been maintained in an ongoing electronic database. Approval of a retrospective chart review has been obtained from the institutional review board of The UTMDACC (Houston, TX).

For the present study, baseline neuropsychological evaluation data of primary brain tumor patients were pooled from the archival database. Patients with primary brain tumors, who did not have prior history of surgery, chemotherapy, or radiation therapy, were selected. Initial sample consisted of 400 patients. From this sample, patients with primary brain tumors predominantly in the cortical regions were selected. Patients were excluded if they had multifocal, bilateral tumors, a prior history of radiation or chemotherapy for a systemic cancer, or fewer than 8 years of education. After these exclusionary criteria were applied, the final sample was comprised of 260 patients.

Measures:

In the archival database, neurocognitive test data, clinician report measure, and self-report symptom measure data of all patients are readily available. For the CTB and the CTBE, scores from the select measures were used (see Table 1). The composition of the comprehensive clinical battery is also presented in Table 1.

Procedure:

Neurocognitive Test Data:

The neuropsychological test scores in the electronic database were converted to normative scores using published normative data based on age, and on education, gender, and handedness, where appropriate. Certain normative systems were chosen based on previous research on psychometric properties and practical suitability of those normative systems. See Table 33 for further details. After obtaining normative data, all standardized scores were converted into *z*-scores. In order to keep the direction of interpretation consistent across component subtests, *z*-scores of TMTA, TMTB, Peg-D, and Peg-ND were multiplied by -1 such that higher z-scores on these measures now indicate superior performance on the neurocognitive measure, and vice versa. Standard scores on the component tests of the CTB and the CTBE were assessed for outliers. Interquartile range (IQR, Q_3 - Q_1) was used to identify outliers operationalized as z-scores that were greater (Q_3 + 1.5*IQR) or less (Q_1 – 1.5*IQR) than 1.5 times the IQR for each component variable (Ghasemi & Zahediasl, 2012, Cousineau & Chartier, 2010, Peat & Barton 2005), which were then trimmed (n = 28).

NCF6Z was then calculated by adding *z*-scores of the six CTB subtests (see Table 2) and dividing the total by 6, such that:

$$NCF6Z = \frac{HVLTRTRSD + HVLTRDRSD + HVLTDRECOGSD + COWASD + TMTASD + TMTBSD}{6}$$

For the CTBE, the BD *z*-score was added to the 6 original scores. The total was then divided by 7, thus generating NCF7Z. Thus,

$$NCF7Z = \frac{HVLTRTRSD + HVLTRDRSD + HVLTDRECOGSD + COWASD + TMTASD + TMTBSD + BDSD}{7}$$

BD scores were not available for 11 patients from the final sample (n = 260). As a result, sample size for NCF7Z was 249.

For the comprehensive clinical battery, *z*-scores of all neurocognitive tests administered were summed. This sum was divided by the total number of outcome scores to obtain an average *z*-score that was called NCFTOTALZ. In the present sample, n = 115 patients were administered the Boston Naming Test (BNT), n = 130 were administered the Visual Naming Test from the Multilingual Aphasia Examination, and n = 15 were not administered any naming test. Patients who received BNT were included in the final sample that constituted NCFTOTALZ. Thus, final sample size for the comprehensive battery was 96.

Scores on self-report measures of symptoms were also available in the database. Given that population norms are available for the STAI, scores on STAI-T and -S were converted to standard scores. For the BDI-II, raw scores were used for further analyses.

Tumor Characteristics, Tumor-related Epilepsy, and Clinician-Report Measures of Patient Functional Status:

Data on tumor grade, laterality, and caudality were available in the electronic database. Similarly, information on seizure history was also included in the database and was readily available for use. Two clinician-report measures of patient functional status, the KPS and the FIM (1987) were included in the archival database. The patients' physician

determined the KPS, while the neuropsychologist determined the FIM scores based on direct patient observation and interview.

Clinician Ratings of Neurocognitive Impairment:

Clinical ratings of neurocognitive impairment were obtained from two experienced clinical neuropsychologists. In order to avoid bias, clinicians were blinded to patient identity and had access only to the relevant patient demographics, which included: study ID, dates of birth, age, gender, handedness, years of education, seizure status, and information on tumor characteristics. Two clinical ratings were obtained from both clinicians for each patient: domain specific ratings of impairment and global impairment status.

Domain specific ratings of impairment:

For each neurocognitive domain (see Table 3), clinicians were requested to provide ratings of impairment on a 9-point scale (from 1: above average to 9: severely impaired). A cut-off score of 5 was assigned to indicate definite mild impairment, while a score of 4 indicated borderline neurocognitive status.

Table 3. Neurocognitive Measures Classified by Domains for Clinician Ratings

Domain	Neurocognitive Measures
Attention	DS
Executive Function	TMTB
	COWA
	Sim
Learning	HVLT-R TR
Memory	HVLT-R DR
	HVLT-R DRECOG
Visual-Spatial Performance	BD
Motor Dexterity	Peg-D/Peg-ND
	(most impaired performance)
Processing Speed	DSym
	TMTA
Language	BNT
	VNT
	Token Test
Gross Motor Function	Grip-D/Grip-ND
	(most impaired performance)

Global Impairment status:

Patients whose clinical ratings exhibited impairment (domain ratings of ≥ 5) in more than one neurocognitive domain were given NCF status of "global impairment." Thus, global impairment status was provided on a binary scale (global cognitive impairment present or global cognitive impairment absent). For those cases where the global impairment status ratings were discrepant, we obtained ratings from a third clinical neuropsychologist, which were then used as final ratings for those discrepant cases.

See appendix A for detailed guidelines provided to the clinicians in order to assign domain-specific and global ratings of impairment.

Statistical analyses:

Statistical analyses were performed using PASW Statistics, release version 18.0.0 for Windows (2009, SPSS Inc., Chicago, IL), MedCalc for Windows version 12.2.1 (2012, MedCalc Software Mariakerke, Belgium), and MS Excel for Windows (2010, Microsoft Corporation, Redmund, WA).

Distributions of composite scores were assessed for normality using the graphical (histogram, boxplots) and numerical (Kolmogorov-Smirnov test with Lillefors correction, skewness, and kurtosis) methods.

Pearson product moment correlation coefficients were used to assess the concordance of each of NCF6Z and NCF7Z with NCFTOTALZ. To examine the criterion and concurrent validity related evidence for the CTB, Spearman's rank order correlations (ρ) were calculated between NCF6Z and the KPS and the FIM.

Independent-samples *t*-tests and unbiased effect sizes (*d*; Hedges & Olkin, 1985) were calculated to evaluate the effects of tumor grade, laterality, and caudality as well as seizure history on NCF6Z. The effect of tumor laterality on NCF7Z was also examined using independent-samples *t*-tests and unbiased effect sizes. Additional post-hoc analyses were conducted in some cases.

Blinded clinician ratings of domain specific impairments and global impairment status of all patients were obtained from two board certified clinicians. We calculated weighted kappa index (w_i) to assess the inter-rater agreement between the domain specific ratings. For 60 patients, global impairment status ratings were discrepant between two clinicians. For patients who were given discrepant ratings (n = 60), ratings of domain

specific impairment and global impairment status were obtained from a third independent clinician and used as the final ratings.

In order to assess diagnostic accuracy of NCF6Z and NCF7Z, we calculated sensitivity, specificity, PPV, and NPV (Zweig et al., 1993) at z-scores ≤ -0.50, -0.60, -0.75, -1.00, -1.25, and -1.50 against the standard of clinician ratings of global impairment status. We also generated ROC curves (McNeil et al., 1984, Hanley et al., 1983) for each of NCF6Z and NCF7Z to determine the ideal cutpoints at which the scores would exhibit optimal diagnostic accuracy.

Results

Sample Characteristics:

The average age of the sample was 45.7 years. Half the patients (50.0%) were male and 88.8% were right handed. Average years of education were 14.97. The most common ethnic background was Caucasian (91.2%). About two-third (65.4%) of the patients were diagnosed with high grade glioma. Among high grade glioma cases, 54.7% patients were diagnosed with glioblastoma multiforme (GBM), 43.5% with anaplastic astrocytoma (AA), and 1.7% with anaplastic oligodendroglioma (AO). Tumor was located in either the frontal or the temporal lobes among 89.2% of the patients (see Table 4).

Table 4. Demographic and Clinical Characteristics of the Sample

Variable	n = 260
Age (Years)	
Mean (SD)	45.65 (13.95)
Range	18-77
Education (Years)	
Mean (SD)	14.97 (2.43)
Range	9-20
Gender	
Female	50.0%
Male	50.0%
Race	
Caucasian	91.2%
African American	1.2%
Hispanic	4.6%
Asian	1.2%
Other	1.9%
Handedness	
Right	88.8%
Left	8.8%
Mixed	2.3%
Grade	
Low Grade Glioma	34.6%
Anaplastic Astrocytoma	28.5%
Anaplastic Oligodendroglioma	1.2%
Glioblastoma	35.8%
Tumor Location	
Frontal Lobe	48.8%
Temporal Lobe	40.4%
Parietal Lobe	10.4%
Occipital Lobe	0.4%
Tumor Laterality	
Left	65.6%
Right	34.4%
Tumor Caudality	
Anterior	89.2%
Posterior	10.8%
Tumor-related Epilepsy	
None	43.1%
Generalized	21.5%
Partial	30.8%
Other	4.6%

Sample size, means and standard deviations of each of the component tests of the comprehensive battery are represented in Table 5.

Table 5. Performance on component subtests of the comprehensive battery

Measure	Sample Size	Mean	SD	*p* value[1]
DS	260	-0.10	0.94	.076
DSym	260	-0.07	0.91	.239
TMTA	260	-0.01	0.97	.824
TMTB	260	-0.99	1.90	.0001
COWA	260	-0.71	1.17	.0001
SIM	252	0.03	1.01	.663
HVLT-R TR	260	-0.94	1.34	.0001
HVLT-R DR	260	-0.96	1.54	.0001
HVLT-R DRECOG	260	-0.63	1.40	.0001
BD	249	-0.02	0.92	.679
BNT	115	-0.76	1.14	.0001
Token Test	249	0.21	0.95	.001
Peg-D	251	-0.60	1.11	.0001
Peg-ND	244	-0.57	0.98	.0001
Grip-D	250	-0.53	0.91	.0001
Grip-ND	248	-0.43	0.97	.0001

[1] *p* value for one-sample t-test

Table 6. Intercorrelations between measures of the Comprehensive Battery

	TR1	DR1	DRECOG1	COWA	TMTA	TMTB	BD	DS	DSy	Sim	BNT	Token	GripD	GripND	PegD
DR1	0.804														
	<.001														
DRECOG1	0.486	0.566													
	<.001	<.001													
COWA	0.345	0.287	0.207												
	<.001	<.001	.001												
TMTA	0.193	0.215	0.151	0.072											
	.002	<.001	.015	.250											
TMTB	0.382	0.316	0.308	0.213	0.411										
	<.001	<.001	<.001	.001	<.001										
BD	0.314	0.347	0.277	0.157	0.249	0.439									
	<.001	<.001	<.001	.013	<.001	<.001									
DS	0.342	0.272	0.177	0.289	0.120	0.286	0.190								
	<.001	<.001	.004	<.001	.054	<.001	.003								
DSy	0.371	0.328	0.233	0.20	0.406	0.465	0.359	0.282							
	<.001	<.001	<.001	.001	<.001	<.001	<.001	<.001							
Sim	0.361	0.384	0.245	0.249	0.079	0.280	0.350	0.331	0.312						
	<.001	<.001	<.001	<.001	.211	<.001	<.001	<.001	<.001						
BNT	0.284	0.334	0.218	0.263	0.016	0.157	0.234	0.147	0.066	0.385					
	.002	<.001	.02	.005	.862	.093	.014	.116	.483	<.001					
Token	0.277	0.309	0.177	0.222	0.10	0.360	0.308	0.343	0.137	0.263	0.315				
	<.001	<.001	.005	<.001	.114	<.001	<.001	<.001	.031	<.001	<.001				
GripD	0.073	0.090	0.058	0.044	0.166	0.189	0.077	0.054	0.231	0.028	0.178	0.106			
	.253	.155	.358	.487	.009	.003	.238	.398	<.001	.665	.060	.098			
GripND	0.00	0.010	0.094	0.022	0.059	0.015	0.008	0.039	0.092	0.055	0.154	0.028	0.633		
	.999	.871	.141	.730	.353	.818	.900	.536	.149	.397	.106	.665	<.001		
PegD	0.133	0.140	0.073	0.13	0.269	0.314	0.266	0.139	0.366	0.20	0.112	0.192	0.20	-0.26	

	.035	.026	.252	.039	<.001	<.001	<.001	.027	<.001	.002	.243	.003	.002	.681	
PegND	.039	0.066	0.066	0.067	0.296	0.238	0.189	0.128	0.229	0.035	0.097	0.143	0.148	0.187	0.509
	.548	.304	.302	.299	<.001	<.001	.004	.046	<.001	.596	.316	.029	.022	.004	<.001

Assessment of Normality of NCF6Z, NCF7Z, and NCFTOTALZ:

Distributions of NCF6Z, NCF7Z, NCFTOTALZ, and NCF7ZW were assessed and found to comply with normal distributions.

Validity Related Evidence for the CTB:

Correlation with NCFTOTALZ:

NCF6Z (r = 0.92, $p \leq .0001$) and NCF7Z (r = 0.92, $p \leq .0001$) correlated significantly with NCFTOTALZ. As expected, NCF6Z and NCF7Z correlated significantly also with each other (r = 0.99, $p \leq .0001$), as shown in Table 6.

Table 6. Correlation between NCF6Z, NCF7Z, and NCFTOTALZ

	n	Pearson's r	*p* value
NCF6Z and NCFTOTALZ	96	0.919	≤ .0001
NCF7Z and NCFTOTALZ	96	0.922	≤ .0001
NCF6Z and NCF7Z	260	0.991	≤ .0001

Tumor Grade:

NCF6Z successfully distinguished between low and high grade tumors, in that patients with high grade glioma ($Mean_{NCF6Z}$ = -0.86, SD_{NCF6Z} = 0.98) obtained significantly lower NCF6Z scores ($t_{(258)} = 3.62$, $p \leq .0001$, d = 0.49) compared to patients with low grade gliomas ($Mean_{NCF6Z}$ = -0.42, SD_{NCF6Z} = 0.81). Post-hoc analyses with Tukey's HSD revealed that NCF6Z of GBM ($Mean_{NCF6Z}$ = -1.01, SD_{NCF6Z} = 0.99) patients was significantly lower compared low grade glioma patients ($p \leq .0001$), but did not differ significantly from NCF6Z of AA patients ($p \leq .083$). See Table 7 and Table 8.

Similarly, patients with high grade gliomas ($Mean_{NCF7Z}$ = -0.75, SD_{NCF7Z} = 0.90) obtained significantly lower NCF7Z scores ($t_{(247)} = 3.41$, $p \leq .001$, d = 0.47) compared to patients with low grade gliomas ($Mean_{NCF7Z}$ = -0.36, SD_{NCF7Z} = 0.77). Post-hoc analyses revealed that NCF7Z was significantly lower ($p \leq .0001$) for patients diagnosed with GBM ($Mean_{NCF7Z}$ = -0.88, SD_{NCF7Z} = 0.90), when compared with low grade glioma patients. The difference between NCF7Z scores of GBM and AA patients was statistically not significant ($p \leq .118$). See Table 7 and Table 8.

Table 7. Difference in Mean NCF6Z Scores (and NCF7Z) Based on Tumor Grade

	Tumor Grade	Mean (SD)	t value	df	*p* value	Cohen's d
NCF6Z	LGG[1]	-0.42 (0.81)	3.62	258	.0001	0.49
	HGG[2]	-0.86 (0.98)				
NCF7Z	LGG[1]	-0.36 (0.77)	3.41	247	.001	0.47
	HGG[2]	-0.75 (0.90)				

[1] Low Grade Glioma
[2] High Grade Glioma

Table 8. Further Assessment of Difference in Mean NCF6Z Scores (and NCF7Z) Based on Tumor Grade

	Tumor Grade	Mean (SD)	F	*p* value	Post-hoc Comparison of Tumor Grade[b]					
NCF6Z	LGG[1]	-0.42(0.81)	6.37	.0001		AA[a]	d[c]	AO[a]	d[c]	GBM[a]	d[c]
	AA[2]	-0.67(0.92)			LGG	.309	0.29	.975	0.18	.0001	0.65
	AO[3]	-0.65(1.58)			AA			1.00	0.02	.083	0.36
	GBM[4]	-1.01(0.99)			AO					.906	0.27
NCF7Z	LGG[1]	-0.36(0.77)	5.60	.001		AA[a]	d[c]	AO[a]	d[c]	GBM[a]	d[c]
	AA[2]	-0.58(0.87)			LGG	.386	0.27	.947	0.24	.0001	0.62
	AO[3]	-0.63(1.39)			AA			1.00	0.04	.118	0.34
	GBM[4]	-0.88(0.90)			AO					.959	0.21

[1] Low Grade Glioma
[2] Anaplastic Astrocytoma
[3] Anaplastic Oligodendroglioma
[4]Glioblastoma Multiforme
[a] *p* value
[b]Tukey's HSD was used for post-hoc analyses
[c]d refers to Cohen's d

Tumor Laterality:

Patients with left hemisphere tumor location ($Mean_{NCF6Z}$ = -0.83, SD_{NCF6Z} = 0.99) obtained significantly lower NCF6Z scores ($t_{(212)} = 2.86$, $p \leq .005$, d = 0.36) compared to those with right hemisphere tumor location ($Mean_{NCF6Z}$ = -0.49, SD_{NCF6Z} = 0.81). Similarly, patients with left hemisphere tumor location ($Mean_{NCF7Z}$ = -0.73, SD_{NCF7Z} = 0.93) obtained significantly lower NCF7Z ($t_{(214)} = 3.03$, $p \leq .003$, d = 0.38) compared to patients with right hemisphere tumor location ($Mean_{NCF7Z}$ = -0.41, SD_{NCF7Z} = 0.73). See Table 9.

Table 9. Difference in Mean NCF6Z Scores (and NCF7Z Scores) Based on Tumor Laterality

	Tumor Laterality	Mean (SD)	t value	df	*p* value	Cohen's d
NCF6Z	Left	-0.82 (0.99)	2.86	212	0.005	0.36
	Right	-0.49 (0.81)				
NCF7Z	Left	-0.73 (0.93)	3.03	214	0.003	0.38
	Right	-0.41 (0.73)				

Analyses of component scores of NCF7Z revealed that HVLT-R-TR ($t_{(257)} = 2.48$, $p \leq .014$, d = 0.33), HVLT-R-DR ($t_{(242)} = 3.66$, $p \leq .0001$, d = 0.46), and COWA ($t_{(257)} = 3.02$, $p \leq .003$, d = 0.41) scores were significantly poorer among patients with left sided tumor location. On the other hand, HVLT-R-DRECOG ($t_{(257)} = 0.66$, $p \leq .509$, d = 0.09), TMTA ($t_{(257)} = 1.71$, $p \leq .089$, d = 0.22), TMTB ($t_{(257)} = 0.42$, $p \leq .672$, d = 0.06), and BD ($t_{(204)} = 0.56$, $p \leq .577$, d = 0.08) did not differ significantly between right and left tumor locations. See Table 10. Furthermore, a one-sample t-test revealed that BD scores (Mean = -0.02, SD = 0.92) in the current sample were not significantly different ($t_{(1)} = 0.41$, $p \leq .679$) from 0.

Table 10. Difference between Component Scores of NCF7Z Based on Tumor Laterality

NCF Measure	Tumor Laterality	Mean (SD)	t value	df	*p* value	Cohen's d
HVLT-R-TR	Left	-1.09 (1.39)	2.48	257	.014	0.33
	Right	-0.66 (1.19)				
HVLT-R-DR	Left	-1.19 (1.68)	3.66	242	.0001	0.46
	Right	-0.54 (1.12)				
HVLT-R-DRECOG	Left	-0.68 (1.45)	0.66	257	.509	0.09
	Right	-0.56 (1.33)				
COWA	Left	-0.87 (1.20)	3.02	257	.003	0.41
	Right	-0.41 (1.05)				
TMTA	Left	-0.08 (0.97)	1.71	257	.089	0.22
	Right	0.13 (0.96)				
TMTB	Left	-0.92 (1.97)	0.42	257	.672	0.06
	Right	-1.03 (1.87)				
BD	Left	-0.05 (0.97)	0.56	204	.577	0.08
	Right	0.02 (0.82)				

Tumor Caudality:

NCF6Z ($t_{(258)}$ = 0.37, $p \leq .711$, d = 0.08) and NCF7Z ($t_{(247)}$ = 0.51, $p \leq .612$, d = 0.11) did not differ significantly between anterior and posterior tumor patients (see Table 11).

Table 11. Difference in Mean NCF6Z Scores (and NCF7Z Scores) Based on Tumor Caudality.

	Tumor Caudality	Mean (SD)	t value	df	*p* value	Cohen's d
NCF6Z	Anterior	-0.71 (0.96)	0.37	258	0.711	0.08
	Posterior	-0.64 (0.81)				
NCF7Z	Anterior	-0.62 (0.89)	0.51	247	0.612	0.11
	Posterior	-0.53 (0.78)				

Further Analyses of Relationship between NCF6Z and Tumor Location:

As many as 89.2% of the present sample (n = 232) was diagnosed with primary tumor location in either frontal or temporal lobes. Thus, we assessed if NCF6Z differentiated between the two anterior lobes. NCF6Z was unable to distinguish ($t_{(230)} = 1.63$, $p \leq .105$, d = 0.22) between frontal ($Mean_{NCF6Z} = -0.62$, $SD_{NCF6Z} = 0.90$) and temporal ($Mean_{NCF6Z} = -0.83$, $SD_{NCF6Z} = 1.02$) tumor locations. NCF7Z exhibited similar trend of results (see Table 12).

Table 12. Difference in Mean NCF6Z Based on Anterior Tumor Location

	Tumor Location	Mean (SD)	t value	df	*p* value	Cohen's d
NCF6Z	Frontal	-0.62 (0.90)	1.63	230	.105	0.22
	Temporal	-0.83 (1.02)				
NCF7Z	Frontal	-0.55 (0.83)	1.28	220	.203	0.18
	Temporal	-0.71 (0.96)				

We further assessed the relationship between NCF6Z and laterality of tumor location for anterior tumors using a 2 x 2 between-subjects Analysis of Variance (ANOVA). Main effect for tumor location (frontal versus temporal) was non-significant: $F_{(1, 228)} < 1$. Significant main effect for tumor laterality $F_{(1, 228)} = 8.30$, $p \leq .004$ suggested that among patients with anterior tumor location, patients with left-sided tumors ($Mean_{NCF6Z} = -0.83$, $SD_{NCF6Z} = 0.99$) performed significantly worse than patients with right sided tumors ($Mean_{NCF6Z} = -0.48$, $SD_{NCF6Z} = 0.84$). Lastly, a significant tumor location x tumor laterality interaction effect: $F_{(1, 228)} = 5.25$, $p \leq .023$ suggested that NCF6Z scores of patients with left temporal tumors ($Mean_{NCF6Z} = -1.01$, $SD_{NCF6Z} = 1.04$) were significantly lower than those of patients with right temporal and frontal tumor locations, while patients with right temporal tumor location performed significantly better ($Mean_{NCF6Z} = -0.32$, $SD_{NCF6Z} = 0.78$). See Table 13, Table 14, and Figure 1.

NCF7Z demonstrated a similar pattern of results. Main effect for tumor location (frontal versus temporal) was non-significant: $F < 1$, while that for tumor laterality $F_{(1, 218)} = 10.13$, $p \leq .002$ was significant, suggesting that among patients with anterior tumor location, patients with left-sided tumors ($Mean_{NCF7Z} = -0.75$, $SD_{NCF7Z} = 0.93$) performed significantly worse than patients with right sided tumors ($Mean_{NCF7Z} = -0.39$, $SD_{NCF7Z} = 0.76$). Interaction between tumor location x tumor laterality was significant: $F_{(1, 218)} = 5.97$, $p \leq .015$. Thus, NCF7Z scores of patients with left temporal tumors ($Mean_{NCF7Z} = -0.90$, $SD_{NCF7Z} = 0.99$) were significantly worse relative to scores of patients with right temporal and frontal tumor locations. Patients with right temporal tumor location performed significantly better ($Mean_{NCF7Z} = -0.19$, $SD_{NCF7Z} = 0.63$). See Table 13, Table 14, and Figure 2.

Table 13. Difference in NCF6Z (and NCF7Z) based on Tumor Laterality x Tumor Location

	Source	SS	df	MS	F value	*p* value
NCF6Z	Tumor Location	0.13	1	0.13	0.15	.699
	Tumor Laterality	7.29	1	7.29	8.30	.004
	Location x Laterality	4.61	1	4.61	5.25	.023
	Error	200.25	228	0.88		
	Total	212.71	231			
NCF7Z	Tumor Location	6.26	1	6.26	0.00	.993
	Tumor Laterality	7.64	1	7.64	10.13	.002
	Location x Laterality	4.50	1	4.50	5.97	.015
	Error	164.33	218	0.75		
	Total	175.85	221			

Table 14. Mean NCF6Z and Mean NCF7Z based on Tumor Laterality in Anterior Tumor Location

			Mean	SD
NCF6Z	Frontal	Right	-0.57	0.86
		Left	-0.65	0.93
	Temporal	Right	-0.32	0.78
		Left	-1.01	1.04
NCF7Z	Frontal	Right	-0.50	0.81
		Left	-0.59	0.86
	Temporal	Right	-0.19	0.63
		Left	-0.90	0.99

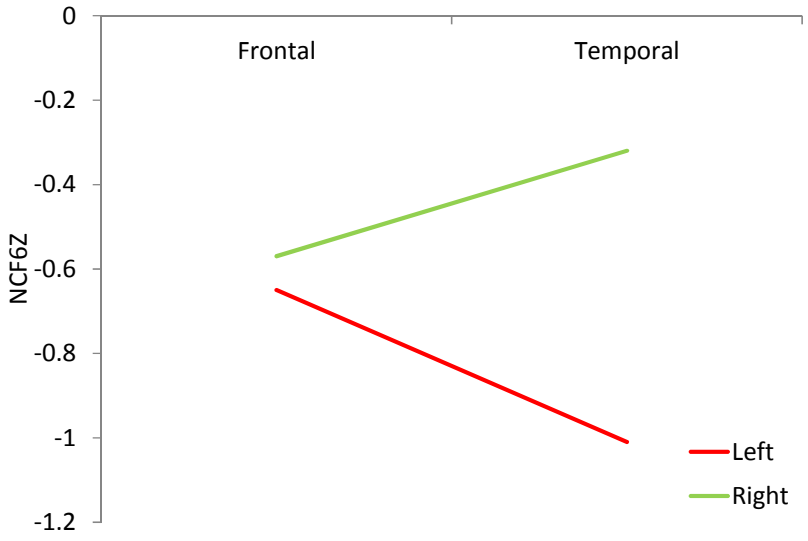

Figure 1. Tumor Location x Tumor Laterality Interaction: NCF6Z

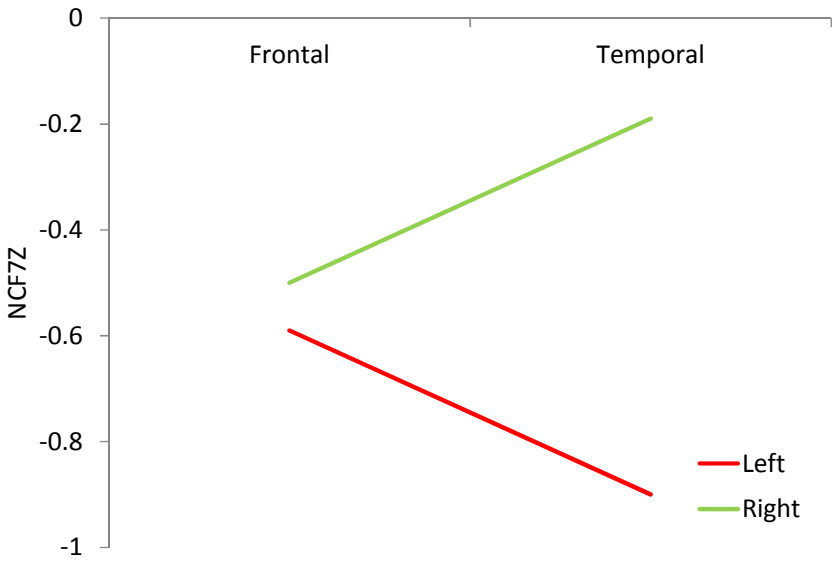

Figure 2. Tumor Location x Tumor Laterality Interaction: NCF7Z

Tumor-related Epilepsy:

NCF6Z scores of patients with tumor-related epilepsy ($Mean_{NCF6Z}$ = -0.67, SD_{NCF6Z} = 0.92) did not differ significantly ($t_{(258)}$ = 0.75, $p \leq .456$, d = 0.09) from patients without tumor-related epilepsy ($Mean_{NCF6Z}$ = -0.76, SD_{NCF6Z} = 0.98). Similarly, NCF7Z did not differ significantly ($t_{(247)}$ = 0.58, $p \leq .564$, d = 0.08) between patients with tumor-related epilepsy ($Mean_{NCF7Z}$ = -0.58, SD_{NCF7Z} = 0.87) and patients without tumor-related epilepsy (MeanNCF7Z = -0.65, SDNCF7Z = 0.89). See Table 15.

Table 15. Difference in Mean NCF6Z Scores (and NCF7Z Scores) Based on Tumor-related Epilepsy

	TrE[1]	Mean (SD)	t value	df	*p* value	Cohen's d
NCF6Z	Present	-0.67 (0.92)	-0.75	258	.456	0.09
	Absent	-0.76 (0.98)				
NCF7Z	Present	-0.58 (0.88)	-0.58	247	.564	0.08
	Absent	-0.65 (0.89)				

[1] Tumor-related Epilepsy

Given that tumor-related epilepsy is more common among low grade glioma patients (Rosati, Tomassini, Pollo, et al., 2009, Klein et al., 2003), we then assessed a subgroup of patients diagnosed with low grade glioma (n = 90). Among patients with low grade glioma, NCF6Z ($t_{(88)}$ = 2.46, $p \leq .016$, d = 0.53) was significantly lower in patients with tumor-related epilepsy ($Mean_{NCF6Z}$ = -0.60, SD_{NCF6Z} = 0.86) compared to patients without tumor-related epilepsy ($Mean_{NCF6Z}$ = -0.19, SD_{NCF6Z} = 0.68). Post-hoc analyses with Tukey's HSD showed that patients with partial seizures obtained significantly lower NCF6Z ($p \leq .025$) compared to patients without tumor-related epilepsy (see Table 16, Table 17, and Table 18). NCF6Z did not differ (t = 0.73, $p \leq .466$) among GBM patients with tumor-related epilepsy compared to GBM patients without tumor-related epilepsy.

NCF7Z demonstrated a similar trend in that patients with low grade glioma (n = 87), in that NCF7Z was significantly lower ($t_{(85)}$ = 2.28, $p \leq .025$, d = 0.50) in patients with tumor-related epilepsy ($Mean_{NCF7Z}$ = -0.52, SD_{NCF7Z} = 0.83) compared to patients without tumor-related epilepsy ($Mean_{NCF7Z}$ = -0.15, SD_{NCF7Z} = 0.65). Post-hoc analyses with Tukey's HSD showed that patients with partial seizures obtained significantly lower NCF7Z ($p \leq$

.031) compared to patients without tumor-related epilepsy. NCF7Z did not differ (t = 0.51, $p \leq .609$) among GBM patients with tumor-related epilepsy compared to GBM patients without tumor-related epilepsy (see Table 16, Table 17, and Table 18).

Table 16. Further Assessment of Difference in Mean NCF6Z Scores (and NCF7Z Scores) Based on Tumor-related Epilepsy among Low Grade Glioma Patients

	TrE[1]	Mean (SD)	t value	df	*p* value	Cohen's d
NCF6Z	Present	-0.60 (0.86)	2.46	88	.016	0.53
n = 90	Absent	-0.19 (0.68)				
NCF7Z	Present	-0.52 (0.83)	2.28	85	.025	0.50
n = 87	Absent	-0.15 (0.65)				

[1] Tumor-related Epilepsy

Table 17. Difference in Mean NCF6Z Scores (and NCF7Z Scores) Based on Tumor-related Epilepsy among Low Grade Glioma Patients

	TrE[1]	Mean (SD)	F	*p* value
NCF6Z	Absent	-0.19 (0.68)	3.15	.029
	Partial	-0.75 (0.88)		
	General	-0.50 (0.75)		
	Other	-0.04 (1.15)		
NCF7Z	Absent	-0.15 (0.65)	3.14	.030
	Partial	-0.67 (0.82)		
	General	-0.41 (0.76)		
	Other	0.08 (1.05)		

[1] Tumor-related Epilepsy

Table 18. Post-hoc Comparison[1] of Difference in Mean NCF6Z (and NCF7Z Scores) Based on Tumor-related Epilepsy among Low Grade Glioma Patients

	TrE[2]	Partial	d[4]	General	d[4]	Other	d[4]
NCF6Z	Absent	.025	0.71	.504	0.43	.984	0.16
	Partial			.733	0.31	.341	0.69
	General					.713	0.47
NCF7Z	Absent	.031	0.71	.651	0.37	.932	0.26
	Partial			.683	0.33	.240	0.80
	General					.639	0.53

[1] Tumor-related Epilepsy
[2] Tukey's HSD was used for post-hoc analyses
[3] *p* value
[4] d refers to Cohen's d

Clinician-report measures:

Distributions of KPS and FIM were not normal. Thus, Spearman's rank order correlations (ρ) were used to derive coefficients of association. NCF6Z significantly correlated with KPS ($\rho = 0.39$, $p \leq .0001$) and the FIM ($\rho = 0.46$, $p \leq .0001$). NCF7Z also correlated significantly with both KPS ($\rho = 0.39$, $p \leq .0001$) and FIM ($\rho = 0.47$, $p \leq .0001$). See Table 19.

.

Self-reported measures of Affective Distress:

BDI-II and STAI-T were not normally distributed. Distribution of STAI-S was normal. BDI-II, STAI-T, and STAI-S did not correlate significantly with NCF6Z. NCF7Z also demonstrated a similar pattern of results. See Table 19.

Further analyses showed that as many about 79% of the patients reported only minimal depressive symptomatology. Similarly, 45% patients reported their present anxiety

levels to be average or less, while 84% patients reported that their anxiety levels are generally within the average range or less (see

Table 20).

Table 19. Correlation between NCF6Z (and NCF7Z Scores) and KPS, FIM, BDI-II, STAI-S, and STAI-T

		n	Correlation Coefficient	*p* value
NCF6Z	KPS	221	0.389[1]	.0001
	FIM	254	0.463[1]	.0001
	BDI-II	126	-0.131[1]	.144
	STAI-S	62	-0.202[2]	.115
	STAI-T	62	-0.212[1]	.098
NCF7Z	KPS	213	0.399[1]	.0001
	FIM	245	0.470[1]	.0001
	BDI-II	62	-0.143[1]	.117
	STAI-S	62	-0.204[2]	.112
	STAI-T	122	-0.219[1]	.088

[1] Spearman's ρ
[2] Pearson's Product Moment Correlation Coefficient r

Table 20. Frequency Analysis of BDI-II, STAI-S, and STAI-T

BDI-II[1] (n = 126)		STAI-S[2]		STAI-T[2]	
	% Patients		% Patients		% Patients
Minimal	78.6	High	54.8	High	16.1
Mild	12.7	Low	45.2	Low	83.9
Moderate	7.9				
Severe	0.8				

[1] BDI-II score: 0-13 = Minimal, 14-19 = Mild, 20-28 = Moderate, 29-63 = Severe
[2]High Anxiety was indicated by a score of 40 or more (Spielberger et al., 1970).

Clinician Ratings of Impairment:

Two types of clinician ratings were obtained—domain specific ratings of impairment (obtained on a 9-point scale for each NCF domain, see Table 3 and Appendix A) and global impairment status (based on domain specific ratings and obtained on a binary scale of present / absent, see Appendix A).

Domain Specific Ratings of Impairment:

Inter-rater agreements between domain specific ratings utilizing the scores of the 260 patients in the final sample were obtained using weighted kappa index with linear weights (Kraemer, Periyakoil, & Noda, 2002). Very good strength of agreement (Altman, 1991) was observed between ratings provided by both clinicians for the domains of learning (w_i = 0.97), memory (w_i = 0.88), attention (w_i = 0.98), language (w_i = 0.88), visuospatial functioning (w_i = 0.99), gross motor functioning (w_i = 0.94), and fine motor dexterity (w_i = 0.91). For the domains speed of processing (w_i = 0.67) and executive function (w_i = 0.65), strength of agreement was good (Altman, 1991). See Table 21.

Table 21. Inter-rater Agreement for Domain Specific Ratings of NCF Status

Cognitive Domain	n	Weighted Kappa Index	95% Confidence Interval
Learning	260	0.974	0.962 to 0.978
Memory	260	0.876	0.850 to 0.903
Processing Speed	260	0.673	0.597 to 0.749
Attention	260	0.982	0.960 to 1.000
Executive Function	260	0.649	0.601 to 0.698
Language	255	0.876	0.842 to 0.909
Visuospatial Function	253	0.991	0.978 to 1.000
Gross Motor Function	256	0.938	0.91 to 0.960
Fine Motor Dexterity	258	0.912	0.878 to 0.946

Inter-rater differences between domain specific ratings were non-significant for all NCF domains except processing speed ($t_{(260)} = 4.87$, $p \leq .0001$, d = 0.20), memory ($t_{(260)} = 2.56$, $p \leq .011$, d = 0.04), gross motor function ($t_{(256)} = 2.44$, $p \leq .015$, d = 0. 03), and fine motor dexterity ($t_{(258)} = 2.31$, $p \leq .022$, d = 0.04). However, all effect sizes of differences between domain-specific clinical ratings were small. See Table 22.

Table 22. Inter-rater Comparison of Mean Domain Specific Ratings of NCF Status

Cognitive Domain		Mean (SD)	t-value	df	*p* value	Cohen's d
Leaning	Clinician 1	3.67 (2.35)	1.15	260	.252	0.01
	Clinician 2	3.69 (2.39)				
Memory	Clinician 1	4.11(2.73)	2.56	260	.011	0.04
	Clinician 2	4.00 (2.46)				
Processing Speed	Clinician 1	2.08 (1.23)	4.87	260	.0001	0.20
	Clinician 2	2.32 (1.21)				
Attention	Clinician 1	2.15 (0.98)	0.90	260	.367	0.02
	Clinician 2	2.13 (0.96)				
Executive Function	Clinician 1	3.82 (2.82)	0.62	260	.539	0.02
	Clinician 2	3.87 (2.16)				
Language	Clinician 1	2.84 (2.05)	1.56	255	.120	0.02
	Clinician 2	2.89 (1.95)				
Visuospatial Function	Clinician 1	2.21 (1.50)	1.34	253	.180	0.16
	Clinician 2	2.01 (0.98)				
Gross Motor Function	Clinician 1	3.09 (1.75)	2.44	256	.015	0.03
	Clinician 2	3.14 (1.71)				
Fine Motor Dexterity	Clinician 1	3.46 (1.98)	2.31	258	.022	0.04
	Clinician 2	3.55 (2.02)				

Global Impairment Status:

Strength of agreement was moderate ($\kappa = 0.53$) between the global impairment status ratings provided by both clinicians (Landis & Koch, 1977). For 77% of the cases, the clinicians provided concordant ratings. This rate of agreement was consistent with that found in other studies using two or more raters (for example, Woods et al., 2004). Nevertheless, ratings of global impairment were discrepant for as many as 23% cases and questions may be raised about the reliability of clinician ratings and impact on assessment of diagnostic accuracy of NCF6Z and NCF7Z.

However, other studies (for example, Woods et al., 2004, Heaton et al., 1994a) have reported usefulness of this method in identifying subtle and diverse forms of cognitive impairment, which is commonly observed in primary brain tumor population. Additionally, for cases (n = 60), where the two clinicians had provided discrepant ratings, we obtained domain specific and global ratings of impairment from a third independent clinician, which then were used as the final ratings for those patients.

In this sample, global impairment was present in 51% of patients, while 49% of patients were "not impaired." Global impairment status was then used for assessment of diagnostic accuracy of NCF6Z and NCF7Z.

Assessment of Diagnostic Accuracy of NCF6Z and NCF7Z:

The cut-off point of z-score ≤ -0.50 has been previously used as an indicator of impairment in research studies (Carey et al., 2004a), because it is roughly equivalent to an average score of -1.0 on approximately one-half of the component measures. Using the cutpoint ≤ -0.50, we assessed sensitivity (0.86), specificity (0.78), PPV (0.80), and NPV (0.85) of NCF6Z. Diagnostic accuracy values of NCF7Z at ≤ -0.50 were as follows: sensitivity (0.84), specificity (0.83), PPV (0.84), NPV(0.83). See Table 23.

Performance at other cutpoints:

Diagnostic accuracy of NCF6Z and NCF7Z was further calculated at cutpoints ≤ -0.60, -0.75, -1.00, -1.25, and -1.50. At cutpoint ≤ -0. 60, sensitivity of NCF6Z was 0.86, sensitivity was 0.82, PPV was 0.83, and NPV was 0.85. At cutpoint ≤ -0. 60, diagnostic accuracy of NCF7Z was as follows: sensitivity = 0.88, specificity = 0.81, PPV = 0.83, NPV =

0.86. At the value of ≤ -1.50, both NCF6Z and NCF7Z became highly specific tests (specificity$_{NCF6Z}$ = 0.99 and specificity$_{NCF7Z}$ = 1.00) with high PPV (PPV$_{NCF6Z}$ = 0.99 and PPV$_{NCF7Z}$ = 1.00), although their sensitivity and NPV values were low (sensitivity$_{NCF6Z}$ = 0.35, sensitivity$_{NCF7Z}$ = 0.30, NPV$_{NCF6Z}$ = 0.59, and NPV$_{NCF7Z}$ = 0.57). See Table 23.

Table 23. Diagnostic Accuracy of NCF6Z and NCF7Z at -0.50, -0.60, -0.75, -1.00, -1.25, and -1.50

	Sensitivity	Specificity	PPV	NPV
NCF6Z ≤ -0.50[a]	0.86	0.78	0.80	0.85
NCF7Z ≤ -0.50[a]	0.84	0.83	0.84	0.83
NCF6Z ≤ -0.60	0.86	0.82	0.83	0.85
NCF7Z ≤ -0.60	0.88	0.81	0.83	0.86
NCF6Z ≤ -0.75	0.73	0.87	0.85	0.75
NCF7Z ≤ -0.75	0.72	0.89	0.88	0.75
NCF6Z ≤ -1.00	0.62	0.91	0.88	0.69
NCF7Z ≤ -1.00	0.55	0.94	0.91	0.66
NCF6Z ≤ -1.25	0.47	0.97	0.94	0.64
NCF7Z ≤ -1.25	0.43	0.98	0.96	0.62
NCF6Z ≤ -1.50	0.35	0.99	0.98	0.59
NCF7Z ≤ -1.50	0.30	1.00	1.00	0.57

[a] Cutpoint commonly used in research studies

ROC Curve Analysis:

An ROC curve was then generated for NCF6Z (see Figure 3), which revealed that the ability of NCF6Z to accurately identify patients with NCF impairment as such was significantly above the chance level (AUC = 89.5%, z = 20.69, $p < 0.0001$). Based on the Youden Index (*J* statistic), NCF6Z = -0.6257 was established as an ideal cutpoint that would afford optimal sensitivity (0.85), specificity (0.83), PPV (0.84), and NPV (0.84). See Table 24.

Another ROC curve was generated for NCF7Z (see Figure 4) and demonstrated that NCF7Z was also able to identify patients with NCF impairments significantly better than random guessing (AUC = 89.6%, z = 20.16, $p < 0.0001$). The Youden Index identified NCF7Z = -0.445 as the optimal cutpoint, at which diagnostic accuracy of NCF7Z was found to be as follows: sensitivity (0.88), specificity (0.81), PPV (0.83), NPV (0.86). See Table 24.

An ROC curve for NCFTOTALZ (see Figure 5) demonstrated that NCFTOTALZ performed better than chance (AUC = 87.5%, z = 10.44, $p < 0.0001$). The Youden Index identified NCF7Z = -0.5063 as the optimal cutpoint, at which diagnostic accuracy of NCF7Z was: sensitivity (0.73), specificity (0.93), PPV (0.92), NPV (0.76). See Table 24.

Table 24. ROC analyses of NCF6Z, NCF7Z, and NCFTOTALZ

	AUC	*p* value	Optimal cutpoint[+]	Sensitivity	Specificity	PPV	NPV
NCF6Z	89.5%	<0.0001	-0.6257	0.85	0.83	0.84	0.84
NCF7Z	89.6%	<0.0001	-0.445	0.88	0.81	0.83	0.86
NCFTOTALZ	87.5%	<0.0001	-0.5063	0.73	0.93	0.92	0.76

[+]Based on highest Youden Index

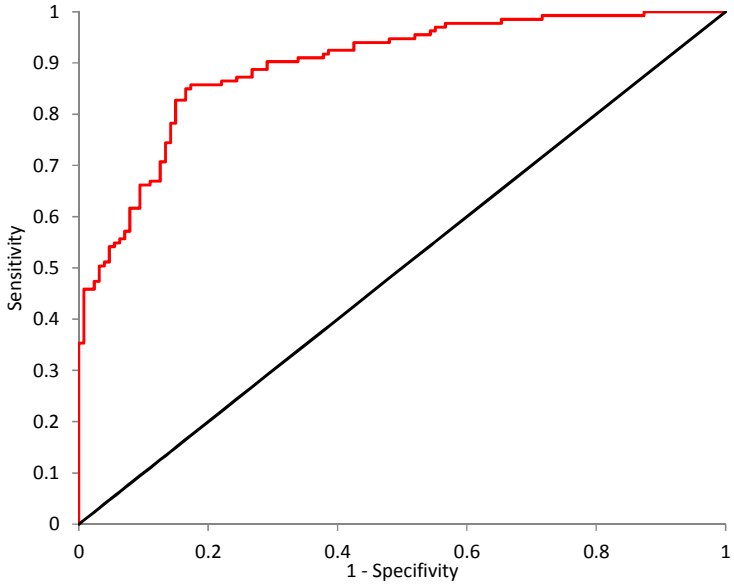

Figure 3. Receiver Operating Characteristic Curve for NCF6Z

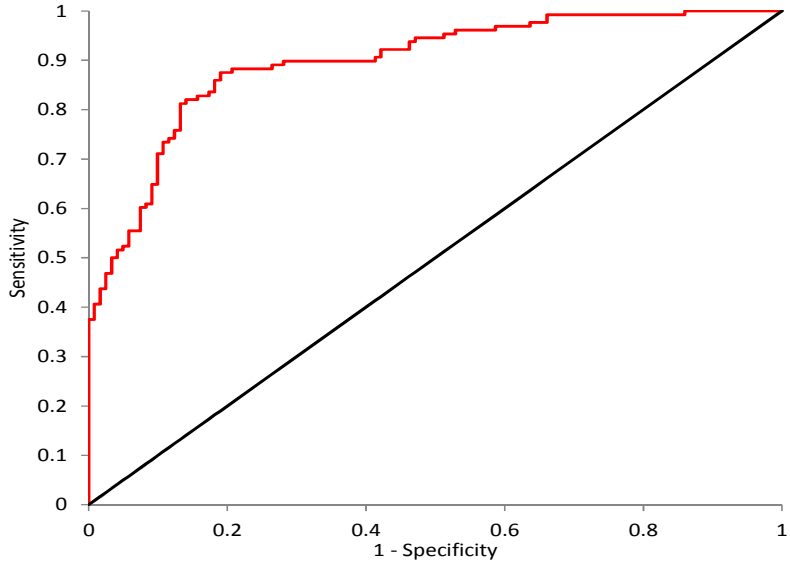

Figure 4. Receiver Operating Characteristic Curve for NCF7Z

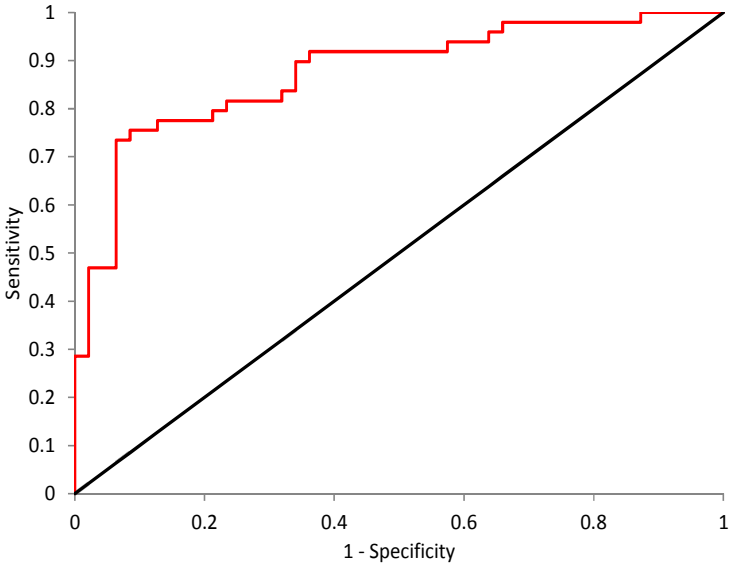

Figure 5. Receiver Operating Characteristic Curve for NCFTOTALZ

Comparison of ROC curves for NCF6Z, NCF7Z, and NCFTOTALZ:

ROC curves for NCF6Z and NCF7Z were compared with each other and with the ROC curve for NCFTOTALZ to examine which composite score offered the best representation of the diagnostic accuracy of the measures it summarized. NCF6Z, NCF7Z, and NCFTOTALZ were significantly superior compared to chance in identifying impaired patients as such. The difference between the AUCs of NCF6Z and NCF7Z was 0.00096 (z-score = 1.08, $p \leq .281$), which was statistically not significant. Difference between AUCs of NCF6Z and NCFTOTALZ as well as that between AUCs of NCF7Z and NCFTOTAL Z was not significant (see Table 25).

Additionally, at their respective *J* values (NCF6Z = -0.6257 and NCF7Z = -0.445), sensitivity, specificity, PPV, and NPV of NCF6Z and NCF7Z were found to be equivalent (see Table 24). Overall, ROC curves for NCF6Z, NCF7Z, and NCFTOTALZ provided

equivalent representation of the diagnostic accuracy of the CTB and the CTBE respectively (see Figure 6).

Table 25. Comparison between AUCs of NCF6Z, NCF7Z, and NCFTOTALZ

	z score	*p* value
NCF6Z and NCF7Z	1.08	.281
NCF6Z and NCFTOTALZ	0.15	.880
NCF7Z and NCFTOTALZ	0.19	.846

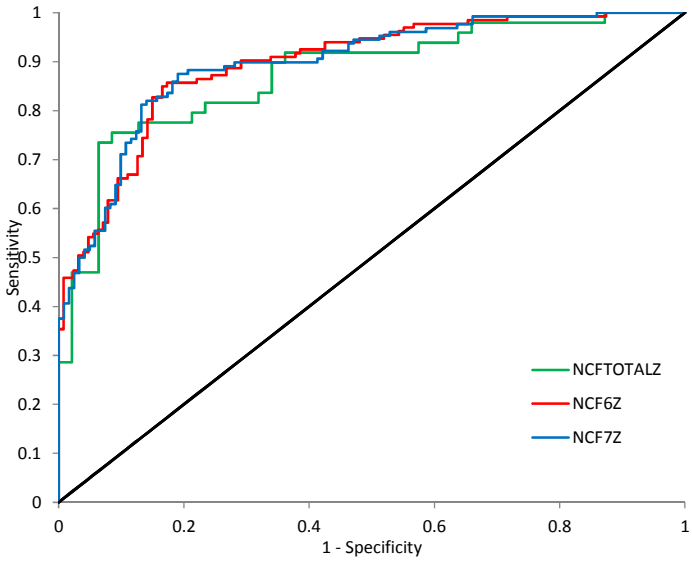

Figure 6. Comparison of Receiver Operating Characteristic Curves for NCF6Z, NCF7Z, and NCFTOTALZ

Additional Analyses:

Another weighted composite score was generated by obtaining the arithmetic mean of NCF6Z and BD ($\frac{(NCF6Z+BD)}{2}$). This composite score was termed NCF7ZW, and gave increased weight to BD (a weight equal to that of the other six measures combined). Validity and diagnostic accuracy of NCF7ZW were assessed.

Assessment of Validity of NCF7ZW:

Distributions of NCF7ZW were assessed and found to conform with the normal distribution. NCF7ZW correlated significantly with NCF6Z (r = 0.86, $p \leq .0001$), NCF7Z (r = 0.92, $p \leq .0001$), and NCFTOTALZ (r = 0.85, $p \leq .0001$). See Table 26.

Table 26. Correlations between NCF7ZW and NCF6Z, NCF7Z, NCFTOTALZ

	n	Pearson's r	*p* value
NCF6Z	249	0.859	≤ .0001
NCF7Z	249	0.919	≤ .0001
NCFTOTALZ	96	0.849	≤ .0001

Consistent with NCF6Z, NCF7ZW was able to successfully distinguish ($t_{(247)}$ = 2.99, $p \leq .003$, d =0.41) between low grade ($Mean_{NCF7ZW}$ = -0.16, SD_{NCF7ZW}= 0.76) and high grade glioma ($Mean_{NCF7ZW}$ = -0.48, SD_{NCF7ZW} = 0.80). See Table 27. Furthermore, post-hoc analyses with Tukey's HSD revealed that NCF7ZW was significantly lower ($p \leq .001$) for GBM patients ($Mean_{NCF7ZW}$ = -0.62, SD_{NCF7ZW}= 0.76) compared to low grade glioma patients. However, difference between NCF7Z performance of GBM and AA patients ($Mean_{NCF7ZW}$ = -0.29, SD_{NCF7ZW}= 0.82) only approached significance ($p \leq .04$) (Table 28).

Table 27. Differences in NCF7ZW Based on Tumor Grade, Laterality, Caudality, and Tumor-related Epilepsy

		Mean (SD)	t value	df	*p* value	Cohen's d
Tumor Grade	LGG	-0.16 (0.76)	2.99	247	.003	0.41
	HGG	-0.48 (0.80)				
Laterality	Right	-0.23 (0.67)	2.20	212	.029	0.29
	Left	-0.45 (0.85)				
Caudality	Anterior	-0.38 (0.80)	0.50	247	.616	0.11
	Posterior	-0.29 (0.78)				
TrE	Present	-0.36 (0.81)	0.69	247	.492	0.06
	Absent	-0.41 (0.78)				

Table 28. Post-hoc Analyses of Difference in Mean NCF7ZW Scores Based on Tumor Grade

Tumor Grade	Mean (SD)	F	*p* value	Post-hoc Comparison of Tumor Grade[b]						
					AA[a]	d[c]	AO[a]	d[c]	GBM[a]	d[c]
LGG[1]	-0.16(0.76)	5.48	.001	LGG	.771	0.16	.774	0.52	.001	0.61
AA[2]	-0.29(0.82)			AA			.901	0.35	.038	0.42
AO[3]	-0.60(0.93)			AO					1.00	0.02
GBM[4]	-0.62(0.76)									

[1] Low Grade Glioma
[2] Anaplastic Astrocytoma
[3] Anaplastic Oligodendroglioma
[4]Glioblastoma Multiforme
[a] *p* value
[b]Tukey's HSD was used for post-hoc analyses
[c]d refers to Cohen's d

Similar to NCF6Z, there was significant difference ($t_{(212)}$ = 2.20, $p \leq .029$, d = 0.29) between NCF7ZW scores of patients with left hemisphere tumor location ($Mean_{NCF7ZW}$ = -0.45 SD_{NCF7ZW}= 0.85) and right hemisphere tumor location ($Mean_{NCF7ZW}$ = -0.23 SD_{NCF7ZW}= 0.67). However, NCF7ZW did not differ significantly ($t_{(247)}$ = 0.50, $p \leq .616$, d = 0.11) between anterior ($Mean_{NCF7ZW}$ = -0.38 SD_{NCF7ZW}= 0.80) and posterior ($Mean_{NCF7ZW}$ = -0.29 SD_{NCF7ZW}= 0.78) tumor patients. In addition, NCF7ZW did not distinguish significantly ($t_{(247)}$ = 0.69, $.p \leq 492$, d = 0.06) between patients with ($Mean_{NCF7ZW}$ = -0.34 SD_{NCF7ZW}= 0.81) and without ($Mean_{NCF7ZW}$ = -0.41 SD_{NCF7ZW}= 0.78) tumor-related epilepsy, although patients with seizures continued to perform better than patients without seizures (see Table 27).

Lastly, NCF7ZW correlated significantly with FIM (ρ = 0.45, $p \leq .0001$) and KPS (ρ = 0.37, $p \leq .0001$), but not with BDI-II (ρ = -0.15, $p \leq .099$), STAI-T (ρ = -0.23, $p \leq .071$), and STAI-S ((ρ = -0.11, $p \leq .398$). This was consistent with NCF6Z (see Table 29).

Table 29. Correlations between NCF7ZW and KPS, FIM, BDI-II, STAI-S, and STAI-T

	n	Correlation Coefficient	*p* value
KPS	213	0.446[2]	.0001
FIM	245	0.369[2]	.0001
BDI-II	122	-0.15[2]	.099
STAI-S	62	-0.192[1]	.134
STAI-T	62	-0.231[2]	.071

[1] Pearson's Product Moment Correlation Coefficient r
[2] Spearman's ρ

Assessment of Diagnostic Accuracy of NCF7ZW:

At cutpoint NCF7ZW ≤ -0.50, sensitivity value was 0.72, while the NPV was 0.75, which was lower than those values for NCF6Z at ≤ -0.50. An ROC curve was generated for NCF7ZW and demonstrated that NCF7ZW performed better than random guessing (AUC = 85.3%, $p < 0.0001$). At *J* value, NCF7ZW = -0.6274, specificity was 0.92, although NPV was only 0.72 (see Table 30, Table 31, and Figure 7).

Table 30. Diagnostic Accuracy of NCF7ZW at -0.50, -0.60, -0.75, -1.00, -1.25, and -1.50

NCF7ZW	Sensitivity	Specificity	PPV	NPV
≤ -0.50[1]	0.72	0.90	0.88	0.75
≤ -0.60	0.67	0.92	0.90	0.73
≤ -0.75	0.59	0.93	0.89	0.68
-1.00	0.41	0.98	0.96	0.61
-1.25	0.24	0.99	0.97	0.55
-1.50	0.16	1.00	1.00	0.53

[1]Cutpoint commonly used in research studies

Table 31. ROC analyses of NCF7ZW

	AUC	*p* value	Optimal cutpoint[+]	Sensitivity	Specificity	PPV	NPV
NCF7ZW	85.3%	<0.0001	-0.6274	0.66	0.92	0.89	0.72

[+]Based on highest Youden Index

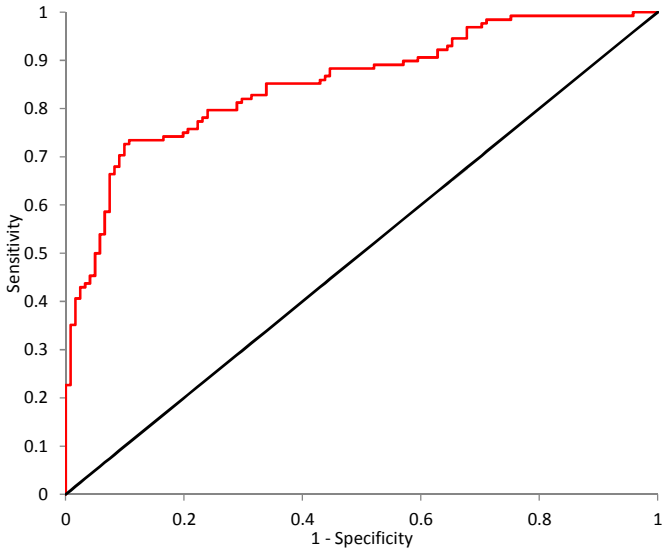

Figure 7. Receiver Operating Characteristic Curve for NCF7ZW

The difference between AUCs of NCF6Z and NCF7ZW was 4.33% (z = 2.43, $p \leq$ .015). See Table 32 and Figure 8.

Table 32. Comparison of AUCs of NCF6Z and NCF7ZW

	Difference between AUCs	z score	*p* value
NCF6Z and NCF7ZW	4.33%	2.43	.015
NCF7Z and NCF7ZW	4.24%	2.97	.003

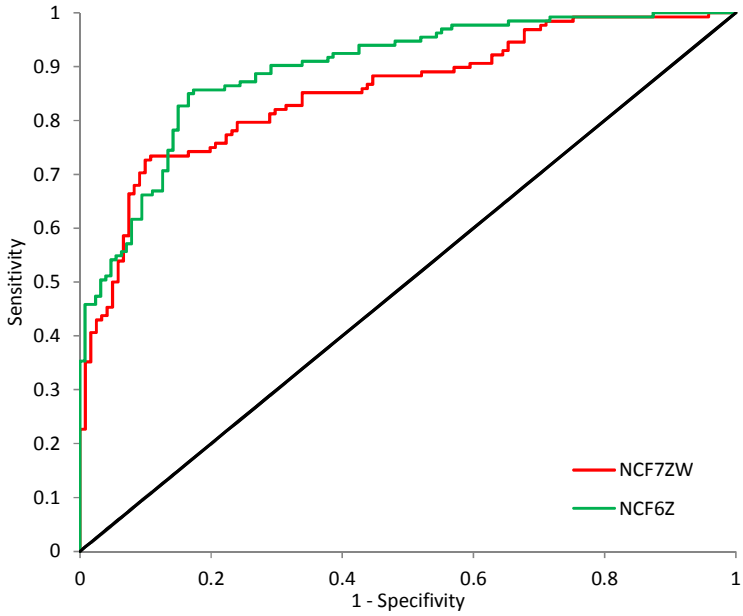

Figure 8. Comparison of Receiver Operating Characteristic Curves for NCF6Z and NCF7ZW

Discussion

This study investigated the validity related evidence and assessed the diagnostic accuracy of the CTB and the CTBE, that are originally derived from a Comprehensive Battery, by generating composite scores NCF6Z and NCF7Z, respectively. Several validation criteria were used to evaluate the validity of the CTB and the CTBE. Consistent with expectations, NCF6Z and NCF7Z correlated significantly with NCFTOTALZ (H1).

We further assessed additional evidence for criterion related validity of NCF6Z and NCF7Z. Patients with low grade glioma obtained significantly lower NCF6Z scores compared with patients with higher grade glioma (GBM, AA and AO). Specifically, patients with GBM performed significantly more poorly than patients diagnosed with low grade

glioma. Thus, the hypothesis that patients with high grade tumors would receive lower NCF6Z scores when compared to those with low grade tumors (H2) was supported. Similarly, NCF7Z scores were significantly lower among patients with high grade tumors relative to patients with low grade glioma.

Consistent with the Hypothesis 3, patients with left hemisphere tumors obtained significantly lower NCF6Z compared with right hemisphere tumor patients. However, patients with left hemisphere tumors also obtained significantly lower NCF7Z relative to right hemisphere tumor patients. Thus, Hypothesis 4 was not supported. The obtained pattern may be explained by a possibility that the addition of only a single test of right hemisphere function (BD) did not adequately offset the purported verbal bias of the CTB. However, even when scores on BD were weighted (NCF7ZW), performance of patients with left hemisphere tumor location remained significantly lower compared to patients with right tumor location. Right hemisphere is considered to be relatively "silent" by some (Kolb & Wishaw, 2003), which may explain performance difference based on tumor location. However, it is also possible that NCF is distributed over cerebral networks (Heimans et al., 2012, Anderson, Damasio, & Tranel, 1990) more than commonly appreciated, and NCF6Z and NCF7Z, in fact, adequately assess patients with heterogeneous lesion locations. Examination of component scores of NCF7Z demonstrated that HVLT-R-TR, HVLT-R-DR, and COWA scores were significantly lower among patients with left sided tumor location, while BD, TMTA, TMTB, and HVLT-R-DRECOG did not differentiate between right and left tumor locations.

Particularly, z-scores on BD tended to cluster around the mean (zero). This may be a sample artifact. Given that BD normative data are not adjusted for educational level, in the

present sample of well-educated patients, it is possible that BD scores exhibited reduced sensitivity, such that normatively average visuoconstructional skills in the current sample may, in fact, represent reduced visuoconstructional functioning, at least in some patients. Alternatively, a restricted range of BD z-scores may be explained by normative structure of WAIS-III subtests, which have standard scores with mean = 10 and SD = 3. Overall, given that BD scores in our sample did not differ from 0, rationale behind choosing BD as a test of right hemisphere / posterior function may be questioned. Choice of BD as a measure of visuoconstruction for the CTBE was based on theoretical as well as practical considerations. BD is a subtest of a widely-used intelligence test (WAIS-III) with robust norms (Wechsler, 1997). Test-retest reliability coefficients range between .80 to .88 (Lezak, Howieson, & Loring, 2004). Factor analytic studies have shown that BD loads highly on the Perceptual Organization Factor (Lezak et al., 2004). In addition to having psychometric properties, BD is practically useful. It is a short and easy to administer test that requires relatively simple and easy to obtain test material. As a result, within the comprehensive battery that is usually administered to primary brain tumor patients in clinical settings at UTMDACC, BD was the best available single measure of right hemisphere / posterior function. Therefore, BD was included as the seventh test of the CTBE.

Although Lezak et al. (2004) report several studies assessing clinical population which confirm the association between lower BD performance and right hemisphere lesions, particularly in the parietal lobe (for example, Newcombe, 1969, Warrington, James, & Maciejewski, 1986, and Wilde et al., 2000 in Lezak et al., 2004), various other studies have shown that BD scores may be affected by other lesion locations as well. For example, patients with significant right parietal damage, those with extensive prefrontal damage, as

well as patients with notable loss of cortical neurons, for example diffuse damage that may be seen in patients with Alzheimer's disease are expected to perform poorly on BD (Luria, 1973b, as reported in Lezak et al., 2004). Lezak et al. (2004) further note that patients with lateralized lesions or "split brain" patients also exhibit subnormal BD performance (Geshwind, 1979), although there may be qualitative differences in their performance. Other research shows that patients with acute traumatic injuries to the frontal lobes, Alzheimer's disease, chronic alcohol dependence, as well as subcortical dementia including Huntington's disease, Parkinson's disease, and Multiple Sclerosis have demonstrated impaired BD performance (Lezak et al., 2004). Thus, performance on BD appears to be influenced by various lesion locations as well as multiple disease processes. In the future, more studies may be conducted to assess sensitivity of BD in primary brain tumor patients.

Finally, it is noteworthy that measures assessing NCF domains of attention (TMTA), executive function (TMTB), and recognition memory (HVLT-R-DRECOG) were among the domains that were robust to tumor laterality. Attention, executive function, and memory are highly critical NCF domains for maintenance of functional independence (Okonkwo, Wadley, Griffith, et al., 2006, Royall, Palmer, Chiodo, et al., 2005).

Hypothesis 5 that patients with anterior tumors would obtain significantly lower NCF6Z compared to those with posterior tumors was not supported. As expected, NCF7Z did not significantly differ between patients with anterior and posterior tumors (H6). As before, it is possible that both NCF6Z and NCF7Z failed to adequately assess posterior brain function. However, weighting the score on BD—a test that is expected to assess posterior function—did not yield a significant difference in NCF7ZW of patients with anterior versus posterior brain tumors. One possible explanation would be the widespread nature of cerebral

networks that support the NCF domains constituting NCF6Z and NCF7Z (Heimans et al., 2012, Anderson et al., 1990), such that NCF6Z and NCF7Z may be adequately assessing patients with anterior and posterior lesion locations. This idea is further supported by an observation that both NCF6Z and NCF7Z correlate significantly with NCFTOTALZ, which is based on a comprehensive battery assessing various NCF domains innervated by heterogeneous neuroanatomical locations.

Alternatively, given that only 10.8% patients had brain tumors primarily in parietal or occipital lobes, small sample size (n = 28) of posterior tumor patients may also explain the obtained results. Further analyses of patients with anterior tumor locations revealed that both NCF6Z and NCF7Z were sensitive to tumor laterality within anterior locations, such that patients with left temporal tumors obtained lowest NCF6Z and NCF7Z scores. This may be explained by the "eloquent" nature of the left temporal lobe (Kolb et al., 2003). Another possible explanation is that NCF6Z and NCF7Z heavily load on learning and memory, which may make them selectively sensitive to left temporal tumors. Lastly, in contrast to patients with temporal lobe tumors, patients with frontal lobe tumors tended to exhibit equivalent NCF6Z and NCF7Z scores. This conforms with the idea that among frontal lesions, laterality of the lesion is less influential in the nature and extent of cognitive impairment (Kolb et al., 2003).

Contrary to expectation, patients with a history of tumor-related epilepsy did not obtain significantly lower NCF6Z compared to patients without history of tumor-related epilepsy. Similarly, NCF7Z was not significantly different for patients with tumor-related epilepsy relative to patients without tumor-related epilepsy (H7). However, further analyses revealed that in a subgroup of patients who were diagnosed with low grade glioma, both

NCF6Z and NCF7Z were significantly lower among patients with tumor-related epilepsy. Among patients with GBM, tumor-related epilepsy was not associated with further lowering of NCF performance. Although research on effect of tumor-related epilepsy among GBM patients is relatively sparse, other studies have observed that high tumor grade is more commonly associated with lower NCF performance rather than comorbid or surgical treatment factors among GBM patients (Talacchi, Santini, Savazzi et al., 2011, Miotto et al., 2011, Bosma et al., 2007), which is consistent with the obtained trend.

Consistent with the hypothesis, NCF6Z correlated significantly with KPS and FIM (H8) and NCF7Z also correlated significantly with KPS and FIM (H9), which provided concurrent validity related evidence for the composite scores. In contrast, the composite scores did not correlate with self-report measures of affective distress, which included scores on STAI-S, STAI-T, and BDI-II. Low percentages of the sample reported significant affective distress ($n_{BDI\text{-}II}$ = 21.4%, n_{STAIS} = 54.8%, and n_{STAIT} = 16.1%). Thus, scores tended to cluster in the subclinical ranges, which may explain the obtained pattern. Fewer patients from the present sample reported anxiety compared to a similar patient group assessed by D'Angelo, Mirijello, Leggio, et al. (2008), in which 62.5% pre-surgical brain tumor patients reported elevated state anxiety and 50% reported high trait anxiety assessed by STAI-S and STAIT, respectively. Similar to the present trend, only 9.7% patients reported depression, as measured by Zung Self-rating Depression Scale. It is unclear why the current patient group reported less anxiety. High level of functional independence may be a possible explanation. In the present sample, the mean KPS score was 92.71 (± 7.38), while the mean FIM score was 123.82 (± 4.35).

Diagnostic accuracy of NCF6Z and NCF7Z was assessed using sensitivity, specificity, PPV, and NPV as the indicators of precision with which NCF6Z and NCF7Z identified global NCF status of patients as impaired. At cutpoint = -0.50, which has been previously used as a cutpoint indicating impairment (Carey et al., 2004a), NCF6Z and NCF7Z were found to be sensitive tests with substantial NPVs. This was consistent with the expectation that NCF6Z and NCF7Z would have adequate sensitivity, specificity, PPV, and NPV (H10, and H11).

We further explored diagnostic accuracy of NCF6Z and NCF7Z at various cutpoints. It was observed that NCF6Z and NCF7Z demonstrated high sensitivity and NPV at cutpoints -0.50 and -0.60. This coincided with the results of ROC curve analyses that revealed NCF6Z $\leq$ -0.6257 and NCF7Z $\leq$ -0.445 to be ideal cutpoints, at which NCF6Z and NCF7Z would exhibit optimal sensitivity, specificity, PPV, and NPV.

Comparison of NCF6Z and NCF7Z:

One of the goals of this study is to compare NCF6Z and NCF7Z to assess which composite score provides a more comprehensive and diagnostically accurate representation of the NCF status of patients with primary brain tumors. Both NCF6Z and NCF7Z correlated significantly with NCFTOTALZ and each other. Both NCF6Z and NCF7Z were able to distinguish between high grade and low grade tumors. Both NCF6Z and NCF7Z were significantly lower among patients with left hemisphere tumor location. Both composite scores did not differ significantly between anterior and posterior tumor patients. Patients with and without tumor-related epilepsy did not obtain significantly different NCF6Z, nor NCF7Z. Both composite scores correlated significantly with KPS and FIM, while and both

did not correlate with self-report measures of affective distress, namely BDI-II, STAI-S, and STAI-T. Overall, NCF7Z did not outperform NCF6Z in validation studies.

Comparison between diagnostic accuracies of NCF6Z and NCF7Z was based on practical considerations for their use. Both scores were found be significantly superior to random guessing. AUCs of NCF6Z and NCF7Z did not significantly differ from each other. Overall, NCF6Z and NCF7Z were largely equivalent in terms of their diagnostic accuracy.

To the extent that the addition of BD, which was expected to afford right hemisphere exposure to the CTB, did not noticeably enhance either the validity or the diagnostic accuracy of the CTB, NCF7Z was not found to have added value compared to NCF6Z. However, it is possible that the CTBE did not achieve adequate exposure to cognitive functions typically localized to the right hemisphere, given that BD was the only index of right hemisphere function within a composite score that consisted of 6 other NCF scores.

Comparison of NCF6Z and NCF7ZW:

NCF7ZW correlated with NCF6Z, NCF7Z, and NCFTOTALZ. Consistent with NCF6Z, NCF7ZW was significantly lower among patients with high grade glioma relative to low grade glioma patients. In particular, GBM patients obtained significantly lower NCF7ZW.

Similar to NCF6Z, NCF7ZW was significantly lower among patients with left hemisphere tumor location and did not distinguish between patients with anterior versus posterior tumor locations. Thus, weighting the scores on BD to enhance right hemisphere / posterior lobar exposure of the CTBE did not extinguish the severity of NCF impairment

among left hemisphere tumor patients, nor did it enhance the NCF differences between tumor caudalities.

Not unlike NCF6Z, NCF7ZW was not significantly different among patients with tumor-related epilepsy, compared to patients without tumor-related epilepsy. Furthermore, NCF7ZW correlated significantly with KPS and FIM and did not correlate with BDI-II, STAI-S, and STAI-T. This performance was consistent with NCF6Z.

At cutpoint -0.50, sensitivity (0.72) and NPV (0.75) of NCF7ZWwere lower than those values for NCF6Z at ≤ -0.50. Both NCF7ZW and NCF6Z performed better than random guessing, but the AUC of NCF6Z (89.5%) was significantly more (4.33%, z = 2.43, $p \leq .015$) than the AUC of (85.3%).

In summary, validation studies showed that NCF7ZW did not outperform NCF6Z. The optimal cutpoints based on an ROC analysis showed a large reduction in sensitivity and improved specificity with the NCF7ZW, which does not support the use of this approach in studies where sensitivity is preferred over specificity. Comparison of their diagnostic accuracy showed that at the cutpoints we assessed as well as in terms of their respective AUCs, NCF6Z was preferable over NCF7ZW.

In conclusion, weighting the score on BD did not offset the purported verbal bias of the CTB. In addition, validation studies and assessment of diagnostic accuracy demonstrated that NCF7ZW did not outperform NCF6Z. The CTB provided an adequately valid representation of the NCF status of the primary brain tumor patients evaluated during this study, when the composite score NCF6Z was assessed against such criterion related and concurrent validation criteria as tumor grade, tumor laterality, association with NCFTOTALZ, KPS, and FIM. NCF6Z sufficiently summarized patient performance on the

CTB such that at NCF6Z = -0.6257, the CTB demonstrated optimal diagnostic accuracy. NCF7Z and NCF7ZW were not found to have superior validity, nor diagnostic accuracy compared to NCF6Z. Thus, the CTBE is not preferable over the CTB. In the interest of efficiency, time- and cost-effectiveness, and training needs of the examiners and also based upon the validity related evidence and diagnostic accuracy of NCF6Z, the CTB is considered a more practical and psychometrically sound battery for use in research studies of primary brain tumor patients.

Future Directions:

Although our sample size was substantially large, in the interest of maintaining homogeneity and avoiding confounds, we studied only primary brain tumors located in the cortical region. Our sample exhibited left-sided and anterior location referral bias. Thus, the obtained results may not be readily generalizable to other types of brain tumors, systemic cancers, or metastases. Additional validation studies may be necessary.

Similarly, our patient pool consisted of chiefly English speaking, college-educated, Caucasian men and women. To the extent that neuropsychological test performance is amenable to cultural, linguistic, and education-based variations, additional studies may be needed to assess the generalizability of the findings of the present study.

Furthermore, it may be of interest to assess if other measures of non-verbal NCF domains add to the psychometric or diagnostic value of the CTB. Lastly, the present study assessed psychometric characteristics of the CTB and NCF6Z among untreated primary brain tumor patients evaluated at the baseline. It may be interesting to assess the ability of NCF6Z

to capture changes in NCF over time, following practice effects, various interventional strategies, and occurrence of other comorbidities.

Conclusion

During the present investigation, a composite score based on unweighted average individual standard scores of the CTB (NCF6Z) was successfully used to summarize the CTB performance. The CTB, which consists of short, easy-to-administer, standardized measures of attention, executive function, learning, memory, and processing speed, demonstrated adequate validity related evidence for use among primary brain tumor patients, when assessed against such concurrent and criterion related factors such as comprehensive NCF battery performance, tumor characteristics, and clinician- report measures of functional status. At cutpoint -0.6257, NCF6Z exhibited optimal diagnostic accuracy.

Addition of a measure of visuospatial functioning (BD) did not augment either validity or diagnostic accuracy of the original CTB, even when the scores were weighted to offset the purported verbal bias of the CTB. In conclusion, the present study provided evidence that NCF6Z is a valid and diagnostically accurate measure of NCF among primary brain tumor patients. The CTB, which is a brief, useful, and easy-to-administer battery, may be used in large scale research studies involving primary brain tumor patients.

Appendix A

Clinician Rating Guidelines:

1. The dataset consists of primary brain tumor patients with supratentorial tumor locations and without prior history of chemotherapy, radiation, or surgery. Please rate each patient's cognitive status based on your clinical experience.
2. The key to interpret demographics:
   a. Sex: m = 1, f = 2
   b. Race: C = 1, AA = 2, H = 3, Asian = 4, Other = 5
   c. Handedness: r = 1, l = 2, mixed = 3
   d. Laterality: r = 1, l = 2

   All patients are pre-treatment, so all treatment indicators will always be 0.
3. The attached document contains patient's baseline scores on cognitive testing that have been converted to demographically corrected scores based on appropriate norms (listed in Table 33). These individual test scores will be classified by neurocognitive domains (see Table 3).
4. For each domain, please provide your ratings of impairment on a 9-point scale. A rating of 1 would signify "above average" performance, while a rating of 9 would mean "severely impaired performance." A cut-off score of 5 on a domain would indicate definite mild impairment in that domain, while a score of 4 would indicate borderline neurocognitive status in that domain.
5. For domains with multiple individual subtests, for example processing speed and EF, you may average the ratings of individual tests to come up with the domain rating, or you may

subjectively assess performance on individual tests to suggest an overall domain rating. If data are missing on a test, please base your ratings on the remaining tests assessing that domain. For tests of motor functioning, please use the most impaired score while assigning the rating.

6. For ease of data entry, please use the excel file provided to you.
7. Patients, whose clinical ratings exhibit impairment (domain ratings of ≥ 5) in more than one neurocognitive domain, would receive neurocognitive status of "global impairment." Please indicate this status as Y (global impairment present) or N (absent) in the corresponding column of the excel file.
8. These ratings will be obtained from two clinicians. For those cases where the ratings are discrepant, ratingsbl from a third clinician will be obtained.

Table 33. Normative Data for NCF measures

Neurocognitive Measure	Normative data
WAIS-III subtests	Wechsler (1997)
HVLT-R	Benedict et al. (1998)
TMT A and B	Tombaugh (2004)
COWA	Ruff et al. (1996)
BNT, PEG, and GRIP	Heaton et al. (2004)
Token Test and Visual Naming	Benton et al. (1994)

References

Anderson, S.W., Damasio, H., Tranel, D. (1990). Neuropsychological impairments associated with lesions caused by tumor or stroke. *Archives of Neurology, 47*(4), 397-405.

Anderson, S.I., Taylor, R., Whittle, I.R. (1999). Mood disorders in patients after treatment for primary intracranial tumours. *British Journal of Neurosurgery,13*, 480-485.

Altman, D.G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

Benedict, R.H.B., Schretlen, D., Groninger, L., Brandt, J. (1998). Hopkins Verbal Learning Test – Revised: Normative Data and Analysis of Inter-Form and Test-Retest Reliability. *The Clinical Neuropsychologist, 12* (1), 43-55.

Benton, A.L., Hamsher, K. deS., Sivan, A.B. (1994). *Multilingual Aphasia Examination* (3rd ed.. San Antonio, TX: Psychological Corporation.

Bottiggi, K.A., Chang, J.J., Schmitt, F.A., Avison, M.J., Mootoor, Y., Nath, A., Berger, J.R., Brown, P.D. (2007). The HIV Dementia Scale: predictive power in mild dementia and HAART. *Journal of Neurological Sciences, 260*(2), 11-15.

Bosma, I., Vos, M.J., Heimans, J.J., Taphoorn, M.J.B., Aaronson, N.K., Postma, T.J., van der Ploeg, H.M., Muller, M., Vandertop, W.P., Slotman, B.J., Klein, M. (2007). The course of neurocognitive functioning in high-grade glioma patients. *Neuro-Oncology, 9* (1), 53-62.

Buckner, J.C., O'Fallon, J.R., Iturria, N.L., O'Neill, B.P., Brown, C.A., Scheithauer, B.W., Dinapoli, R.P., Arusell, R.M., Curran, W.J., Abrams, R., Shaw, E.G. (2004). Importance of baseline mini mental state examination as a prognostic factor for

patients with low-grade glioma. *International Journal of Radiation Oncology, Biology, Physics, 59*, 117–125.

Caine, C., Mehta, M.P., Laack, N.N., Gondi, V. (2012). Cognitive function testing in adult brain tumor trials: lessons from a comprehensive review. *Expert Review of Anticancer Therapy, 12* (5), 655-567.

Carey, C.L., Woods, S.P., Rippeth, J.D., Gonzalez, R., Moore, D.J., Marcotte, T.D., Grant, I., Heaton, R.K. (2004a). Initial validation of a screening battery for the detection of HIV-associated cognitive impairment. *Clinical Neuropsychology, 18 (2),* 234-248.

Carey, C.L., Woods, S.P. Gonzalez, R., Conover, E., Marcotte, T.D., Grant, I., Heaton, R. K., (2004b). Predictive validity of global deficit scores in detecting neuropsychologieal impairment in HIV infection. *Journal of Clinical and Experimental Neiiropsychology, 26 (3),* 307-319.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 38-46.

Cohen, J.A., Cutter, G.R., Fischer, J.S., Goodman, A.D., Heidenreich, F.R., Jak, A.J., Kniker, J.E., Kooijmans, M.F., Lull, J.M., Sandrock, A.W., Simon, J.H., Simonian, N.A., Whitaker, J.N. (2001). Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. Archives of Neurology, 58(6), 961-967.

Cousineau, D., Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research, 3* (1), 58-67.

Cutter, G.R., Baier, M.L., Rudick, R.A. (1999). Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain, 122* (5), 871– 882.

D'Angelo, C., Mirijello, A., Leggio, L., Ferrulli, A., Carotenuto, V., Icolaro, N., Miceli, A., D'Angelo, V., Gasbarrini, G., Addolorato, G. (2008). State and trait anxiety and depression in patients with primary brain tumors before and after surgery: 1-year longitudinal study. *Journal of Neurosurgery, 108* (2), 281-286.

DeAngelis, L. M. (2001). Brain tumors. *The New England Journal of Medicine, 344* (2), 114-123.

Fox, S.W., Mitchell, S.A., Booth-Jones, M. (2006). Cognitive impairment in patients with brain tumors: assessment and intervention in the clinic setting. *Clinical Journal of Oncology Nursing, 10* (2), 169–176.

Garb, H.N., Schramke, C.J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analysis. *Psychological Bulletin, 120,* 140–153.

Ghasemi, A., Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism, 10* (2), 486-489.

Gonzalez, R., Heaton, R.K., Moore, D.J., Letendre, S., Elliott, R., Wolfson, T., Marcotte, T.D., Cherner, M., Rippeth, J., Grant, I., & Group, T.H. (2003). Computerized reaction time battery versus a traditional neuropsychological battery: Detecting HIV-related impairments. *Journal of the International Neuropsychological Society, 9,* 64–71.

Grant, R., Slattery, J. Gregor, A., Whittle, I. R. (1994). Recording neurological impairment in clinical trials of glioma. *Journal of Neuro-Oncology, 19,* 37-49.

Gregor, A., Cull, A., Traynor, E., Stewart, M., Lander, F., Love, S. (1996). Neuropsychometric evaluation of long-term survivors of adult brain tumours:

relationship with tumour and treatment parameters. *Radiotherapy and Oncology, 41*, 55–59.

Hanley, J.A., McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*, 839-843.

Heaton, R.K., Miller, W., Taylor, M.J. & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Odessa, FL: Psychological Assessment Resources.

Heaton, R.K., Grant, I., Butters, N., White, D.A., Kirson, D., Atkinson, J.H., McCutchan, J.A., Taylor, M.J., Kelly, M.D., Ellis, R.J., Wolfson, T., Velin, R., Marcotte, T.D., Hesselink, J.R., Jernigan, T.L., Chandler, J., Wallace, M., Abramson, I. (1995). The HNRC 500 – Neuropsychology of HIV infection at different disease stages. *Journal of the International NeuropsychologicalSociety, 1*, 231–251.

Heaton, R.K., Kirson, D., Velin, R.A., Grant, I. (1994a). The utility of clinical ratings for detecting cognitive change in HIV infection. In I. Grant & A. Martin (Eds.), *Neuropsychology of HIV infection* (pp. 188–206). New York: Oxford University Press.

Heaton, R., Paulsen, J.S., McAdams, L.A., Kuck, J., Zisook, S., Braff, D., Harris, J., & Jeste, D.V. (1994b). Neuropsychological deficits in schizophrenics. Relationship to age, chronicity, and dementia. *Archives of General Psychiatry, 51*, 469–476.

Hedges, L.V., Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Heimans, J.J., Reijneveld, J.C. (2012). Factors affecting the cerebral network in brain tumor patients. *Journal of Neuro-Oncology, (108)* 2, 231-237.

Ingraham, L. J., Aikken, C. B. (1996). An empirical-approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology, 10*, 120–124.

Johnson, D.R., Sawyer, A.M., Meyers, C.A., O'Neill, B.P., Wefel, J.S. (2012). Early measures of cognitive function predict survival in patients with newly diagnosed glioblastoma. *Neuro-Oncology, 14*(6), 808-816.

Kalkers, N.F., Bergers, L., de Groot, V., Lazeron, R.H., van Walderveen, M.A., Uitdehaag, B.M., Polman, C.H., Barkhof, F. (2001). Concurrent validity of the MS Functional Composite using MRI as a biological disease marker. *Neurology, 56* (2), 215-219.

Keefe, R.S., Bilder, R.M., Harvey, P.D., Davis, S.M., Palmer, B.W., Gold, J.M., Meltzer, H.Y., Green, M.F., Miller, D.D., Canive, J.M., Adler, L.W., Manschreck, T.C., Swartz, M., Rosenheck, R., Perkins, D.O., Walker, T.M., Stroup, T.S., McEvoy, J.P., Lieberman, J.A. (2006). Baseline neurocognitive deficits in the CATIE schizophrenia trial. *Neuropsychopharmacology, 31*(9), 2033-2046.

Keefe, R.S., Goldberg, T.E., Harvey, P.D., Gold, J.M., Poe, M.P., Coughenour, L. (2004). The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophrenia Research, 1* (68), 283-297.

Klein, M., Engelberts, N.H., van der Ploeg, H.M. et al. (2003). Epilepsy in low-grade gliomas: the impact on cognitive function and quality of life. *Annals of Neurology, 54* (4), 514–520.

Klein, M., Postma, T.J., Taphoorn, M.J.B., et al. (2003). The prognostic value of cognitive functioning in the survival of patients with high-grade glioma. *Neurology, 61*, 1796-199.

Kolb, B., Wishaw, I.Q. (2003). *Fundamentals of human neuropsychology* (5th ed.). New York, NY: Worth Publishers.

Kraemer, H., Periyakoil, V.S., Noda, A. (2002). Kappa coefficients in medical research. *Statistics in Medicine, 21* (14), 2109-2129.

Lageman, S.K., Cerhan, J.H., Locke, D.E., Anderson, S.K., Wu, W., Brown, P.D. (2010). Comparing neuropsychological tasks to optimize brief cognitive batteries for brain tumor clinical trials. *Journal of Neurooncology, 96(2)*, 271-276.

Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics 33* (1), 159–174.

Lezak, M.D., Howieson, D.B., Loring, D.W. (2004). Neuropsychological Assessment (4th ed.). New York: Oxford University Press.

Litofsky, N.S., Farace, E., Anderson, F., Jr., Meyers, C.A., Huang, W., Laws, E.R., Jr., & Glioma Outcomes Project investigators. (2004). Depression in patients with high-grade glioma: Results of the Glioma Outcomes Project. Neurosurgery, 54, 358–367.

Marciniak, C.M., Sliwa, J.A., Spill, G. et al. (1996). Functional outcome following rehabilitation of the cancer patient. *Archives of Physical Medicine and Rehabilitation, 77*, 54-57.

McNeil., B.J., Hanley J.A. (1984). Statistical approaches to the analysis of Receiver Operating Characteristic (ROC) Curves. *Medical Decision Making, 4* (2), 137-150.

Meyers, C.A., Brown, P.D. (2006). Role and relevance of neurocognitive assessment in clinical trials of patients with CNS tumors. *Journal of Clinical Oncology, 24(8),* 1305–1309.

Meyers, C.A., Hess, K.R. (2003). Multifaceted end points in brain tumor clinical trials: cognitive deterioration precedes MRI progression. *Neuro-oncology, 5*, 89–95.

Meyers, C.A., Hess, K.R., Yung, W.K., Levin, V.A. (2000). Cognitive function as a predictor of survival in patients with recurrent malignant glioma. *Journal of Clinical Oncology,18(3),* 646-50.

Meyers, C.A., Smith, J.A., Bezjak, A., Mehta, M.P., Liebmann, J., Illidge, T., Kunkler, I., Caudrelier, J.M., Eisenberg, P.D., Meerwaldt, J., Siemers, R., Carrie, C., Gaspar, L.E., Curran, W., Phan, S.C., Miller, R.A., Renschler, M.F. (2004). Neurocognitive function and progression in patients with brain metastases treated with whole-brain radiation and motexafin gadolinium: results of a randomized phase III trial. *Journal of Clinical Oncology*, *22*(1), 157-165.

Miller L.S., Rohling, M.L. (2001). A Statistical interpretive method for neuropsychological test data. *Neuropsychology Review, 11*, 143-169.

Miller, D.M., Rudick. R.A., Cutter, G., Baier, M., Fischer, J.S. (2000). Clinical significance of the multiple sclerosis functional composite: relationship to patient-reported quality of life. *Archives of Neurology, 57* (9), 1319-1324.

Miotto, E.C., Junior A.S., Silva C.C., Cabrera H.N., Machado M.A., Benute G.R., Lucia M.C., Scaff M., Teixeira M.J. (2011). Cognitive impairments in patients with low grade gliomas and high grade gliomas. *Arquivos de Neuro-psiquiatria, 69*(4), 596-601.

Okonkwo, O.C., Wadley, V.G., Griffith, H.R., Ball, K., Marson, D.C. (2006). Cognitive correlates of financial abilities in mild cognitive impairment. *Journal of American Geriatric Society, 54* (11), 1745-1750.

Peat, J., Barton, B. (2005). *Medical Statistics: A guide to data analysis and critical appraisal*. Ames: Blackwell Publishing

Rajneesh, K. F. & Binder, D.K. (2009). Tumor-associated epilepsy. *Neurosurgical Focus, 27*(2), 1-4.

Rasmussen, T. (1957). Surgery of epilepsy associated with brain tumors. *Advances in Neurology, 8*, 227–239.

Réthelyi, J.M., Czobor, P, Polgár, P., Mersich, B., Bálint, S. et al. (2012). General and domain-specific neurocognitive impairments in deficit and non-deficit schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience, 262* (2), 107-115.

Rosati, A., Tomassini, A., Pollo, B., Ambrosi, C., Schwarz, A., Padovani, A., Bonetti, B.J. (2009). Epilepsy in cerebral glioma: timing of appearance and histological correlations. *Neuro-oncology, 93* (3), 395-400.

Royall, D.R., Palmer, R., Chiodo, L.K. et al. (2005). Executive control mediates memory's association with change in instrumental activities of daily living: The Freedom House Study. *Journal of American Geriatric Society, 53*, 11-17.

Ruff, R.M., Light, R.H., Parker, S.B., Levin, H.S. (1996). Benton Controlled Oral Word Association Test: reliability and updated norms. *Archives of Clinical Neuropsychoogy, 11*(4), 329-38.

Spielberger, C.D., Gorsuch, R., Luthene, R.E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Tabarés-Seisdedos, R., Balanzá-Martínez, V., Sánchez-Moreno, J., Martinez-Aran, A., Salazar-Fraile, J., Selva-Vera, G., Rubio, C., Mata, I., Gómez-Beneyto, M., Vieta, E. (2008). Neurocognitive and clinical predictors of functional outcome in patients with schizophrenia and bipolar I disorder at one-year follow-up. *Journal of Affective Disorders, 109* (3), 286-299.

Talacchi, A., Santini, B., Savazzi, S., Gerosa, M. (2011). Cognitive effects of tumour and surgical treatment in glioma patients. *Journal of Neuro-Oncology, 103* (3), 541-549.

Taphoorn, M.J., Klein M. (2004). Cognitive deficits in adult patients with brain tumors. *Lancet Neurology, 3(3),* 159-168.

Tombaugh, N. (2004). Trail Making Test A and B: normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19*(2), 203-214.

Wechsler, D. (1997). *WAIS-III Administration and Scoring Manual*. San Antonio, TX: Harcourt Brace & Co.

Wefel, J.S., Kayl, A.E., Meyers, C.A. (2004). Neuropsychological dysfunction associated with cancer and cancer therapies: a conceptual review of an emerging target. *British Journal of Cancer, 90(9),* 1691-1696.

Weitzner, M.A., Meyers, C.A. (1997). Cognitive functioning and quality of life in malignant glioma patients: A review of the literature. *Psycho-Oncology, 6*, 169–177.

Weitzner, M. A., Meyers, C. A. Byrne, K. (1996). Psychosocial functioning and quality of life in patients with primary brain tumours. *Journal of Neurosurgery, 84*, 29-34.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer, 3*, 32-35.

Zweig, M.H., Campbell, G. (1993). Receiver operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*, 561–577.