

BRIEF COMMUNICATION

Cognitive estimation in traumatic brain injury

NOAH D. SILVERBERG,¹ ROBIN A. HANKS,² AND CHERISSE MCKAY¹

¹Department of Psychology, University of Windsor, Windsor, Ontario, Canada

²Department of Rehabilitation Psychology and Neuropsychology, Rehabilitation Institute of Michigan, and Department of Physical Medicine and Rehabilitation, Wayne State University School of Medicine, Detroit, Michigan

(RECEIVED November 25, 2006; FINAL REVISION March 25, 2007; ACCEPTED March 26, 2007)

Abstract

The present study explores the construct and ecological validity of the Biber Cognitive Estimation Test (BCET) in a traumatic brain injury (TBI) sample. Participants completed the BCET in the course of a neuropsychological evaluation at 1–15 years after injury. BCET scores correlated moderately with other standard measures of executive functioning, and contrary to our hypotheses, at least as high with neuropsychological tests with minimal demands on executive functioning. Moreover, partialing out the portion of BCET variance not attributable to executive functioning markedly attenuated the former correlations. With respect to ecological validity, BCET scores did not predict concurrent functional status, as measured by the Disability Rating Scale. By comparison, standard measures of executive functioning strongly correlated with each other, correlated less strongly with nonexecutive functioning measures, and predicted functional status. In conclusion, unlike standard measures of executive functioning, the BCET demonstrated poor construct and ecological validity in TBI patients. (*JINS*, 2007, 13, 898–902.)

Keywords: Neuropsychological tests, Neurobehavioral manifestations, Head injuries, Closed head injuries, Penetrating, Cognition, Judgment

INTRODUCTION

Cognitive estimation (CE) tasks involve asking questions for which there are no readily knowable answers, such as “How long does it take for fresh milk to go sour in the refrigerator?” Examinees must, therefore, provide a reasonable estimate rather than merely retrieve factual information from memory. A response is typically scored as “incorrect” or deviant if it diverges considerably from a range of estimates provided by a normative sample. This paradigm was originally developed to model a common real-world undertaking that heavily taxes aspects of executive functioning (EF) such as working memory and response-monitoring (Shallice & Evans, 1978).

After nearly three decades of research on CE, it remains unclear whether Shallice and Evans’ (1978) task or its subsequent revisions (e.g., Axelrod & Millis, 1994; Brand et al., 2003; Gillespie et al., 2002) do indeed (1) place strong demands on EF and (2) have relevance for real-world func-

tioning. Inconsistent reports of convergent and ecological validity are likely due, at least in part, to a major methodological limitation—there are almost as many versions of the CE task as there are research studies on CE. Most researchers develop their own measure of CE, often changing both the stimuli (i.e., questions) and scoring method. Such modifications are probably not trivial, as suggested by one study showing that correlations between a few different CE tests were remarkably low, in the order of .01 to .15 (Gillespie et al., 2002).

The validity of the CE paradigm in traumatic brain injury (TBI) patients is even less understood, despite that it may hold considerable potential for neuropsychological assessment and rehabilitation planning given the high rate of EF impairment and the pressing need for more ecologically relevant measures in this population (e.g., Sherer et al., 2002). Performance on older versions of the CE task was lower in patients with severe TBI relative to medical controls (Axelrod & Millis, 1994) and contributed to a regression model predicting psychosocial recovery from TBI (Shretlen, 1992), engendering some grounds to believe that the CE paradigm may prove to be a clinically useful adjunct measure in this population.

Correspondence and reprint requests to: Noah D. Silverberg, Department of Psychology, University of Windsor, 173 Chrysler Hall South, 401 Sunset Ave., Windsor, Ontario N9B 3P4, Canada. E-mail: noahsilverberg@hotmail.com

The Biber Cognitive Estimation Test (BCET; Bullard et al., 2004) is the most recent iteration of the CE paradigm, improving on the shortcomings of many previously developed measures, such as better norms, questions without real answers, not providing the units, and equal sampling from multiple domains. Providing some evidence for construct validity, it was found to correlate strongly with a well-established measure of EF (i.e., number of perseverative errors on the Wisconsin Card Sorting Test; Heaton et al., 1993) in samples of schizophrenic patients (Jackson, 2002) and adolescents with pervasive developmental disorders (Liss et al., 2000). In the latter study, this correlation remained strong even after partialing out a measure of verbal intelligence. The BCET was also able to predict vocational status in a subset of patients with schizophrenia following their participation in a rehabilitation program (Jackson, 2002), suggesting that it may also possess some degree of ecological validity. Inspired by these initial findings, the goal of the present study was to examine the BCET's construct and ecological validity in a TBI sample.

A popular method for demonstrating construct validity is to show strong correlations with other measures of the same, or at least an overlapping, construct (convergent validity) and weak correlations with measures of theoretically independent constructs (discriminant validity). With regard to convergent validity, significant associations with EF tests would be less than impressive, given the highly multifactorial nature of the BCET. Good BCET performance probably involves different aspects of EF, including cognitive flexibility, self-monitoring, and working memory, as well as non-EF cognitive abilities, primarily general world knowledge (Axelrod & Millis, 1994; Brand et al., 2003). Statistically partialing out variance on a well-matched control task from the total variance of a multifactorial task with presumably high demands on EF may isolate the "EF component" of that task (Denckla, 1996). This strategy is akin to the clinical practice of considering a relatively preserved performance on Part A of the Trail Making Test (Reitan & Wolfson, 1985) to enhance the interpretability of low scores on Part B as reflecting the impairment of EF (i.e., set-shifting ability). The natural control task for the CE paradigm is one that measures general world knowledge, or semantic memory. The Information subtest of the Wechsler batteries, which requires examinees to produce factual responses to trivia-style questions, exemplifies such a task. The amount of residual variance in the BCET that is shared with EF tests, after partialing out performance on this task, would be a more stringent measure of convergent validity (cf., Liss et al., 2000).

Ecological validity is informed by the correspondence between test performance and quantified real-world functioning (Chaytor & Schmitter-Edgecombe, 2003). Aspects of real-world functioning with relevance to neurocognitive status include work or academic performance, independence with living and activities of daily living, and so on. Although these constructs are notoriously difficult to measure, several psychometrically reliable questionnaires,

observer rating scales, and structured clinical interviews have been developed for this purpose.

Using these procedures, we evaluated the construct and ecological validity of the BCET in TBI patients. We predicted that the BCET would (1) correlate with standard neuropsychological measures of EF, even after its semantic memory component is partialled out; (2) correlate less strongly with neuropsychological measures that do not recruit EF as extensively; and (3) predict functional status above and beyond standard neuropsychological measures of EF.

MATERIALS AND METHODS

Participants

Seventy-seven consecutive participants in the Southeastern Michigan Traumatic Brain Injury Systems program between May 2004 and December 2005 were included in this study. To be eligible, participants had to meet at least one of the following inclusion criteria: (1) posttraumatic amnesia duration > 24 hours, (2) trauma-related to intracranial neuroimaging abnormalities, and (3) Glasgow Coma Scale score of less than 13 in the emergency department. Therefore, injury severity ranged from mild-complicated to severe. The demographic and clinical characteristics of this sample are presented in Table 1.

Materials and Procedure

Participants completed the BCET as part of a research battery that also included the Information^a and Letter-Number Sequencing subtests of the Wechsler Adult Intelligence Scale, Third Edition (raw score; The Psychological Corporation, 1997), Trail Making Test (time to complete Part B; Reitan & Wolfson, 1985), and Color-Word subtest of the Delis-Kaplan Executive Functioning System (time to complete Interference/Stroop Trial; Delis et al., 2001). The latter three measures were chosen because they enjoy widespread clinical use and are relatively factor-pure measures of working memory, set-shifting, and response inhibition, respectively, and therefore (albeit somewhat crudely) map on the Miyake et al. (2000) three-factor model of EF. The recognition discriminability measure of the California Verbal Learning Test, Second Edition (d-prime; Delis et al., 2000), Judgment of Line Orientation (total correct; Benton et al., 1983), and Grooved Pegboard Test (dominant hand, time to complete; Reitan & Wolfson, 1985) were selected because of their minimal recruitment of EF.

The Disability Rating Scale (DRS; Rappaport et al., 1982) is a widely used measure of functional outcome of TBI,

^aSeveral participants did not complete the Information subtest because of time constraints (it was the last test in a battery with a standardized order of administration). These participants did not differ on the BCET from those who completed both measures [$t(75) = .529, p = .598$]. They also did not differ with respect to age, obtained education level, severity of injury, or time since injury (all $p > .05$). Sample size for each pairwise correlation was never less than 65.

Table 1. Demographic and clinical characteristics of the study sample

Age, years:	Mean (<i>SD</i>)	43.71 (14.66)
	Range	18–80
Education:	Mean (<i>SD</i>)	12.01 (2.24)
	Range	6–18
Gender:	Male/Female	65/12
Race:	African American	53
	Caucasian	21
	Other	3
Glasgow Coma Scale score:	Median	10
	Range	3–15
Loss of consciousness duration, days:	Mean (<i>SD</i>)	3.6 (5.6)
	Range	.5–35
Posttraumatic amnesia duration, days:	Mean (<i>SD</i>)	21.0 (14.9)
	Range	.5–71
Time since injury, years:	Mean (<i>SD</i>)	6.8 (5.2)
	Range	1–15

spanning the continuum from vegetative state to complete independence with instrumental activities of daily living. High scores reflect poorer functional status (i.e., greater disability). It is administered in interview format and has excellent inter-rater reliability, in the range of .97–.98. A review of its reliability and validity is available from The Center for Outcome Measurement in Brain Injury (<http://www.tbims.org/combi/drs/>).

Each response on the BCET was scored as correct if it fell between the 5th and 95th percentile of the normative sample in the derivation study (Bullard et al., 2004) and incorrect if it fell outside of this range or if the examinee produced a nonsensical unit of measurement (e.g., pounds for distance). Total scores were then computed by summing the number of correct items, and so could hypothetically range from 0 to 20. Of note, the BCET was re-scored using the Axelrod & Millis (1994) deviation scoring method; the 2nd, 16th, 84th, and 98th percentiles for the normative sample provided by the test's author (Bullard, personal communication, 2006). This score correlated extremely highly with the BCET total score ($r = -.891$, $p < .01$) and did not change the pattern of results reported below. This study received ethical approval from Wayne State University Institutional Review Board.

RESULTS

The BCET demonstrated moderate internal consistency among its 20 items (Cronbach's $\alpha = .610$). Each of the four domain scores demonstrated weaker internal consistency, ranging from .192 (Weight) to .444 (Time). The mean BCET total score within the sample was 15.54 ($SD = 2.80$), with a range from 9 to 20. Of our sample, 45.5% ($n = 35$) scored below the suggested conservative cutoff score by Bullard

et al. (2004) of 3 *SD* below the normative mean (15.6/20). This finding may indicate that the BCET is sensitive to the effects of TBI, but should be interpreted with caution given the substantial dissimilarity between the demographic characteristics of our sample and the Bullard et al. (2004) comparison group.

All variables involved in the below analyses were first screened for factors that can distort the magnitude of parametric (Pearson product-moment) correlations, including skewness, outliers, restricted ranges, bivariate nonlinearity, and heteroscedasticity. Only one serious violation was detected: the DRS was severely positively skewed. Because using a logarithmic transformation of this variable did not substantially alter the results described below, the untransformed data are reported. A power analysis revealed that our sample size was sufficient to find modest-sized correlations ($>.30$) to be significant ($\alpha = .05$, one-tailed) with adequate ($>80\%$) power.

The correlation matrix of the BCET and other standard measures of EF is shown in Table 2. The BCET correlated modestly but significantly with the standard measures of EF (all $p < .05$), but its median (absolute) correlation (.28) was substantially lower than that between the standard measures (.66). As a more stringent test of convergent validity, we then partialled out an appropriate control task from the BCET and standard measures of EF (except that we unfortunately did not have data on a control task for the Letter-Number Sequencing, such as a forward digit span procedure), and examined their residual shared variance with each other. This should theoretically remove the variance associated with the non-EF processes involved in these tasks (Denckla, 1996). After Information scores were partialled out, the BCET's correlations with the standard measures dropped markedly (median = .12) and were statistically nonsignificant. When Part A of the Trail Making Test was partialled out, Part B correlated .53 ($p < .01$) with Letter-Number Sequencing and .36 ($p < .01$) with the Stroop task. Similarly, when the color naming trial was partialled out,

Table 2. Neuropsychological test correlation matrix

	BCET	LNS	TMT-B	Stroop
BCET	—			
LNS	.255*	—		
TMT-B	-.357**	-.652**	—	
Stroop	-.278*	-.660**	.668**	—
JOLO	.347**	.484**	-.482**	-.476**
GPT-D	-.361**	-.465**	.557**	.496**
Recognition	.465**	.522**	-.531**	-.433**

Note. BCET = Biber Cognitive Estimation Test; LNS = Letter-Number Sequencing subtest of the Wechsler Adult Intelligence Scale, Third Edition; TMT-B = Trail Making Test-Part B; Stroop = Color-Word subtest of the Delis-Kaplan Executive Functioning Systems; JOLO = Judgment of Line Orientation; GPT-D = Grooved Pegboard Test, dominant hand; Recognition = recognition discriminability of the California Verbal Learning Test, Second Edition.

* $p < .05$.

** $p < .01$.

the Stroop task correlated .45 ($p < .01$) with the Trail Making Test–Part B and .43 ($p < .01$) with Letter–Number Sequencing.

With respect to discriminant validity, the BCET actually correlated somewhat higher with the tests that have minimal EF demands (median = .36) than the standard executive functioning measures. The opposite (predicted) trend was seen for the standard measures, which averaged a correlation of .48 with measures of visuospatial perception, fine motor dexterity, and recognition memory. These data are also shown in Table 2.

In terms of its ecological validity, on its own, the BCET did not predict functional status as measured by the DRS ($r = .20, p = .09$). In contrast, Letter–Number Sequencing ($r = -.31, p < .01$), Trail Making Test–Part B ($r = .42, p < .01$), and Stroop ($r = .42, p < .01$) were all significantly associated with the DRS. To determine whether the BCET adds to the ability of neuropsychological test scores to predict functional status, the three standard measures of EF were regressed on DRS scores in the first step of a hierarchical regression. This model was significant [$F(3,58) = 5.84, p < .01, R^2 = .232$]. The BCET total score was then added in a second step. It explained an additional 0% of variance, and accordingly, the change in R^2 was far from significant [$F(1,57) = .00; p = .99$]. In other words, the BCET did not improve the prediction of functional status over and above standard measures, *at all*.

Each of the aforementioned analyses were repeated for the four BCET domain scores (time/duration, quantity, weight, and distance). Similar to the overall BCET score, each domain score correlated moderately with the standard measures of EF but of at least equal magnitude as the non-EF measures. Furthermore, with only one exception ($r_{BCET-time * Trails B} = -.350, p < .05$), all correlations with EF measures fell below statistically significant levels when Information scores were partialled out. Finally, only the time domain score significantly correlated with the DRS ($r = -.28, p < .05$). However, it did not contribute to the prediction of functional status above and beyond the three standard measures of EF [R^2 change = .00, $F(1,57) = .01; p = .94$].

DISCUSSION

The purpose of the present study was to evaluate the construct and ecological validity of the BCET in a traumatic brain injury sample. We hypothesized that cognitive estimation places heavy demands on EF, and as such, should correlate strongly with standard measures of EF while demonstrating relatively weak relationships with neuropsychological measures that minimally involve EF. In contrast with these predictions, the BCET demonstrated poor construct validity in terms of both convergent and discriminant validity. Although it showed modest correlations with standard measures of EF (working memory, set-shifting, and response inhibition), these correlations were strongly attenuated by partialing out the variance associated with the semantic mem-

ory (non-EF) component of the BCET. By comparison, the standard measures of EF were strongly intercorrelated and remained so after their respective non-EF components were partialled out. Of note, our obtained pattern of BCET correlations was at odds with Liss et al. (2000), who reported a very strong relationship between the BCET and a measure of set-shifting ($-.63$), even after a verbal intelligence measure was partialled out. The global cognitive compromise in their pervasive development disorder sample likely explains why the pattern of shared variance differed from our TBI sample, in which EF is disproportionately impaired.

Our second main hypothesis was that cognitive estimation captures a unique aspect of real-world functioning, and so should add to the prediction of functional status over standard neuropsychological measures. This hypothesis was also not supported, as the BCET failed to predict functional status on its own and explained no additional variance in functional status beyond that accounted for by the standard EF measures. Unlike the BCET, the standard EF measures were associated with functional status, which is consistent with previous findings (Hanks et al., 1999), and further supports the clinical use of these tests to predict independence with activities of daily living.

That construct and ecological validity were demonstrated for the standard measures of EF but not the BCET helps to rule out sample-specific factors (e.g., injury severity/chronicity, ethnic composition, etc.)^b or other methodological limitations (e.g., skewed distribution of our outcome measure) as accounting for the latter finding. Rather, circumscribed null findings indicate test-specific shortcomings, at least in a TBI population.

REFERENCES

- Axelrod, B.N. & Millis, S.R. (1994). Preliminary standardization of the cognitive estimation test. *Assessment, 1*, 269–274.
- Benton, A., Hamsher, K. deS., Varney, N.R., & Spreen, O. (1983). *Contributions to neuropsychological assessment*. New York: Oxford University Press.
- Brand, M., Kalbe, E., Fujiwara, E., Huber, M., & Markowitsch, H.J. (2003). Cognitive estimation in patients with probable Alzheimer's disease and alcoholic Korsakoff patients. *Neuropsychologia, 41*, 575–584.
- Bullard, S.E., Fein, D., Gleeson, M.K., Tischer, N., Mapou, R.L., & Kaplan, E. (2004). The Biber Cognitive Estimation Test. *Archives of Clinical Neuropsychology, 19*, 835–846.
- Chaytor, N. & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature

^bDespite this reasoning, we examined the relationships between demographic and clinical characteristics of our sample and BCET performance. Only the correlation with education attainment was significant [$r(76) = .25, p = .03$]. Because of this finding, we divided the sample into subjects who had at least a high school diploma ($n = 41$) and those with less education ($n = 36$) and re-ran all the correlational analyses in each subgroup. The BCET correlated similarly with the standard EF measures, measures with minimal EF demands, and Disability Rating Scale in both groups, indicating that education level did not modify the patterns of shared variance between the BCET and other variables.

- on everyday cognitive skills. *Neuropsychological Review*, *13*, 181–197.
- Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan executive functioning system*. San Antonio, TX: The Psychological Corporation.
- Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (2000). *California Verbal Learning Test-Second Edition: Adult version*. San Antonio, TX: The Psychological Corporation.
- Denckla, M.B. (1996). Research on executive function in a neurodevelopmental context: Application of clinical measures. *Developmental Neuropsychology*, *12*, 5–15.
- Gillespie, D.C., Evans, R.I., Gardener, E.A., & Bowen, A. (2002). Performance of older adults on tests of cognitive estimation. *Journal of Clinical and Experimental Neuropsychology*, *24*, 286–293.
- Hanks, R.A., Rapport, L.J., Millis, S.R., & Deshpande, S.A. (1999). Measures of executive functioning as predictors of functional ability and social integration in a rehabilitation sample. *Archives of Physical Medicine and Rehabilitation*, *80*, 1030–1037.
- Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.C., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and expanded*. Lutz, FL: Psychological Assessment Resources, Inc.
- Jackson, C.T. (2002). *Validation of the Biber Cognitive Estimation Test for person with schizophrenia involved in vocational rehabilitation program*. Dissertation, University of Connecticut, Hartford.
- Liss, M., Fein, D., Bullard, S., & Robins, D. (2000). Cognitive estimation in individuals with pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *30*, 613–618.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., & Howarter, A. (2000). The unity and diversity of executive functions and their contribution to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- The Psychological Corporation. (1997). *Wechsler Adult Intelligence Scale* (3rd Ed.). San Antonio, TX: The Psychological Corporation.
- Rappaport, M., Hall, K.M., Hopkins, K., Belleza, T., & Cope, D.N. (1982). Disability rating scale for severe head trauma: Coma to community. *Archives of Physical Medicine and Rehabilitation*, *63*, 118–123.
- Reitan, R. & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press.
- Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*, 293–303.
- Sherer, M., Novack, T.A., Sander, A.M., Struchen, M.A., Alderson, A., & Thompson, R.N. (2002). Neuropsychological assessment and employment outcome after traumatic brain injury: A review. *Clinical Neuropsychologist*, *16*, 157–178.
- Shretlen, D.J. (1992). Accounting for variance in long-term recovery from traumatic brain injury with executive abilities and injury severity. *Journal of Clinical and Experimental Neuropsychology*, *14*, 77.