2013

Discriminant Construct Validity of ImPACT™: A Companion Study

Arthur C. Maerlender
University of Nebraska-Lincoln, amaerlender2@unl.edu

L. Flashman
Geisel School of Medicine at Dartmouth

A. Kessler
University of Wisconsin School of Medicine

S. Kumbhani
University of Wisconsin School of Medicine

R. Greenwald

See next page for additional authors

Authors
Arthur C. Maerlender, L. Flashman, A. Kessler, S. Kumbhani, R. Greenwald, Dartmouth College Toteson, and T. McAllister

Published in *The Clinical Neuropsychologist* 27:2 (2013), pp. 290-299;
doi: 10.1080/13854046.2012.744098.

Accepted for publication October 22, 2012; published online January 14, 2013.

Discriminant Construct Validity of ImPACT™: A Companion Study

A. Maerlender,[1] L. Flashman,[1] A. Kessler,[2] S. Kumbhani,[3]
R. Greenwald,[4] T. Tosteson,[5] and T. McAllister[1]

1 Department of Psychiatry, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA
2 Department of Psychiatry, University of Wisconsin School of Medicine
3 Private practice
4 Simbex LLC and Thayer School of Engineering, Dartmouth College, Hanover, NH, USA
5 Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

Corresponding author — A. Maerlender

Abstract

In a previous analysis of ImPACT scores relative to traditional neuropsychological tests (NP) and experimental tasks (Maerlender et al., 2010) we demonstrated convergent construct validity for the primary ImPACT test-score composites. A complete analysis of discriminant validity was not undertaken at that time. Here, test scores from the 54 collegiate football and hockey players were re-analyzed to specifically address the discriminant validity of the ImPACT composite scores using a multiply operationalized correlation matrix of multitrait multi-method data. In the method used here, discriminant validity is determined by obtaining non-significant correlations between a target score when correlated with the average of the other trait measurements (multiply-operationalized multitrait- mono-method analysis). Results showed that the ImPACT Verbal Memory (p = .044), Visual Memory (p = .006), and Visual Motor Speed (p = .000) scores were highly correlated with composites of the other scores, while the Reaction Time composite demonstrated adequate discriminant validity (p = .145). In comparison all of the NP composites showed good discrimination (all p-values > .05, except for Reaction Time p = .05). Thus the apparent lack of discriminability between three of four composite scores in this sample raises questions about using ImPACT composite scores to support specific construct-oriented interpretations. Taken together, the discriminant and convergent construct validity properties of ImPACT indicate construct sensitivity, but limited construct specificity.

Keywords: Sports concussion, ImPACT, Contruct validity, Discriminant validity, Neuropsychological testing

Introduction

The use of neuropsychological testing in the assessment of brain injury has a long history. Neuropsychological testing for the assessment and monitoring of sport-related concussion has also been practiced for many years. However, with the advent of computerized test batteries and web-based storage of data, neuropsychological screening tests have become almost ubiquitous in the field of sports medicine. The use of these tests for establishing baseline (pre-injury) function is considered to be an important advancement in rehabilitation and recovery monitoring. By using each athlete's own baseline, changes in post-injury test scores should be more sensitive and specific to the athlete's recovery. Still, studies concerning reliability and validity of these tests are important for understanding their usefulness, and obtaining a reliable baseline assessment with a valid test is an important first step in the assessment process. The fact that these tests are screening tests changes the psychometric requirements somewhat, but does not change the importance of establishing validity and reliability although, for screening tests, sensitivity is typically more important than specificity. With that in mind we undertook a multi-trait, multi-method assessment of the construct validity of the ImPACT test (Immediate Post-Concussive Assessment and Cognitive Test), which is the most widely used sports concussion neuropsychological test in use. The initial analyses documented levels of convergent validity in a sample of college football and hockey players (Maerlender et al., 2010). In this analysis aspects of discriminant validity are explored.

Establishing construct validity is a process. Campbell and Fiske (1959) recommended that construct validity demonstrates both high correlations with tests of supposed similar constructs and low correlations with tests from which it should differ. These processes were described as ''convergent'' and ''discriminant'' validity, respectively. Convergent validity is thus a type of construct validity that examines the degree to which the operationalization of specific constructs or traits (i.e., test scores) are similar to (converge on) test scores that they would be expected to be related to. In contrast, discriminant validity demonstrates that different or unique traits do not correlate with each other. Campbell and Fiske (1959) went on to articulate the process of multi-trait multi-method analysis of construct validity, by which constructs of interest (traits) are measured by multiple means (methods). In this way better construct specification can be obtained while controlling for shared method variance.

Campbell and Fiske (1959, p. 100) also noted that: ''For the justification of novel trait measures, for the validation of test interpretation, or for the establishment of construct validity, discriminant validation as well as convergent validation is required. Tests can be invalidated by too high correlations with other tests from which they were intended to differ.'' They go on to assert that each test employed is a trait-method unit, which contains both trait features, and measurement features that are not specific to the trait content. ''In order to examine discriminant validity, and in order to estimate the relative contributions of trait and method variance, more than one trait as well as more than one method must be employed in the validation process'' (p. 100). Systematic variance among test scores is related to both test response features and trait content features. By utilizing a multi-trait multi-method matrix (MTMM), trait and method features can be more clearly isolated. A similar approach is used in structural equations measurement modeling (SEM) when multiple trait-method units are obtained.

Discriminant validity is an index of difference. That is, an index of the difference between the test/construct of interest and some other test/construct. In this approach the constructs that are assessed within each method are compared (correlated) with the mean of the other constructs to show how different the target construct is from other constructs. For example, if a visual memory tests correlates highly with a verbal memory test, there is not adequate discriminant validity as it is expected that these tests assess different underlying constructs. In the multiple-operationalization the visual memory test is correlated with the mean of verbal memory, processing speed, and reaction time tests. In this example the visual memory test's usefulness for specific identification of visual memory deficits is called into question. Thus, while convergent validity seeks high correlations among similar constructs, discriminant validity expects low correlations between dissimilar constructs. Some of the usefulness of this approach is that calculations combining different traits (heterotrait) measured by the same method (mono-method) highlights the role of (shared) method variance: since the trait information should be different for each composite, strong correlations suggest levels of method variance that blurs the specificity of trait information. Alternatively, high correlations may indicate universal trait representation in the tasks. Of course, not all variance may be method related.

Campbell and Fiske's original conceptualization used visual inspection of the full matrix, making judgments based on similarities and differences among correlations. Heterotrait correlations (both heteromethod and mono-method) were expected to show smaller correlations than mono-method-heterotrait correlations. Cole, Howard, and Maxwell (1981) and Howard, Maerlender, Myer, and Curtain (1992) extended Campbell and Fiske's original strategy by using a multiplyoperationalized approach. Rather than relying on visual inspection of validity ''triangles'' used by Campbell and Fiske (the array of trait by method intercorrelations), Cole and colleagues (1981) demonstrated that combining *z*-scores of trait-method units would allow for examination of relationships in various trait method units.

For a mono-method, multiple trait analysis, the *z*-scores for the separate traits within one method are calculated; the mean of three *z*-scores is then correlated with the *z*-score of the fourth trait. All combinations are calculated and the correlations are inspected for significance and pattern.

Previous studies have assessed aspects of concurrent validity of ImPACT with other traditional neuropsychological tests (Iverson, Franzen, Lovell, & Collins, 2003; Iverson, Lovell, & Collins, 2005; Schatz & Putz, 2006), which generally supported the validity of some ImPACT composites. For example, the Symbol- Digit Modalities test (Smith, 1991) was shown to be strongly correlated with all composites in each of the three studies.

An assessment of construct and concurrent validity was conducted by Allen and Gfeller (2011) utilizing ImPACT and a battery of traditional neuropsychological tests used by the National Football League. Their findings revealed high correlations among many indicators from both batteries; however, non-significant correlations were present as well. Importantly, discriminant validity was not explicitly assessed during that study.

In our 2010 study (Maerlender et al.) we compared paper and pencil neuropsychological battery test scores to ImPACT scores and fMRI task responses to obtain estimates of convergent construct validity in a sample of non-injured college athletes (males). Strong intercorrelations were observed for four of the five ImPACT composites (factors) with the paper and pencil battery. Impulse control was the one ImPACT factor that did not demonstrate adequate convergent construct validity. Working memory and sustained attention from the NP battery did not correlate with any ImPACT scores. Scores from affect rating scales were also correlated with some factors: state anxiety was significantly related to ImPACT processing speed.

As opposed to the other studies in which only a subgroup of tests and ImPACT domain scores were analyzed, we assessed the construct validity of ImPACT domain scores by comparing them to a more extensive battery weighted to assess the broad spectrum of cognitive sequelae commonly associated with mild TBI (mTBI) including measures of visual and verbal memory, working memory, attention, processing speed, fine motor skills, and mood symptoms. Indeed, those findings demonstrated that there were strong mono-trait by hetero-method correlations (similar traits by different methods) between the ImPACT domains and the NP domains. While discriminant validity was mentioned in that article, at that time we noted the likely presence of adequate discriminant validity based on score patterns; a more formal analysis had always been intended.

The present study intended to analyze the discriminant validity of ImPACT factors to extend the analysis of ImPACT construct validity. Due to sample size constraints a multiply-operationalized procedure for analyzing the correlation matrix of *z*-transformed scores was used to demonstrate which ImPACT tests were significantly different from the others as an indicator of discriminant construct validity (Cole et al., 1981).

Method

Participants

For this analysis the same data as considered in Maerlender et al. (2010) were analyzed. A total of 54 collegiate football players were administered pre-season ImPACT computerized tests and a battery of more traditional neuropsychological (paper and pencil NP) tests. While 13 had histories of prior concussions, all were healthy at the time of baseline testing (i.e., none was recovering from a recent concussion). Cases were eliminated that exceeded the validity indicators provided by ImPACT. Due to a change in the larger study (the addition of visual memory tests during the course of the study), only 33 of the athletes were administered the visual memory tests.

Measures

As part of the research protocol all participants completed a comprehensive neuropsychological testing and ImPACT testing. This was completed over one or two sessions. Testing was conducted in a soundproofed, well-ventilated office, and tests were administered by a trained examiner under the supervision of a neuropsychologist (LAF) Participants were encouraged to take breaks as needed. All data were checked by another examiner and the neuropsychologist; ImPACT scores were reviewed by a certified examiner (AM).

ImPACT battery. Version 2.0 of the web-based ImPACT program was administered to all athletes at the start of their regular season. We used the ImPACT domain scores and subtest

Table 1. Composition of domain scores for ImPACT, neuropsychological tests (NP) and experimental tests

Domain	ImPACT Composite scores	NP Composite scores
Verbal Memory	Word Memory test (% correct), Match (hidden symbols scores), Three Letters (total correct).	CVLT (trials 1–5 total; long delay total recognition discriminability).
Visual Memory	Xs and Os (total correct memory), Design Memory (total % correct.	BVMT-R (trial 1, total learning trials 1–3, delayed recall).
Reaction Time	Xs and Os (average correct RT), Symbol Match (average correct RT), Color Match (average correct RT).	CPT (Simple Reaction Time, Vigilance & Distractibility average reaction times).
Visual Motor Speed/ Processing Speed	Xs and Os (total correct interference score), Three Letters (average counted correctly).	DKEFS Trail Making (sum of trials 1–3 and 5), DKEFS Verbal Fluency (sum of Category & Letter Fluency), DKEFS Color Word Interference Test (sum of 4 conditions).

Adapted from Maerlender et al. (2010). BVMT-R=Brief Visual Memory Test, Revised (Benedict, 1997); CVLT=California Verbal learning test, 2nd ed (Delis, Kramer, Kaplan, & Ober, 2000); Gordon, 1986; DKEFS=Delis Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2001); PASAT=Paced Auditory Serial Attention Test, Gronwall, 1977.

scores, which are provided by the ImPACT program. Table 1 outlines the tests that comprise these domains. These composite scores are generated by combining two or three subtest scores from the tests that are administered. The composites are not empirically derived.

Neuropsychological battery. A comprehensive battery of neuropsychological tests, which lasted approximately 2 hours, was administered to all participants. This battery was designed to be sensitive to cognitive functions known to show deficits or changes in the context of TBI. Domain scores were created based on the consensus opinion of three neuropsychologists from the Dartmouth Neuropsychology Laboratory. Each participant's raw test data were *z*-transformed using published normative data. The *z*-scores for each test within a given domain were averaged together to create a domain score. The battery also included selfreport measures of mood (the Beck Depression Inventory, 2nd edition; Beck-II; Beck, Steer, & Brown) and anxiety (Spielberger State-Trait Anxiety Questionnaire; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). In addition, the Word Memory Test (Green & Astner, 1995; Green, 2007; Hunt, Ferrara, Miller, & Macchocchi, 2007) was administered to assess effort.

The test batteries appear in Table 1, and the inter-correlation matrix for ImPACT and NP scores appear in Table 2 (alpha set at .05).

Impulse Control variables were not analyzed for discriminant validity because they demonstrated no convergent construct validity in the previous study.

Table 2. Intercorrelations (*p*-values), for ImPACT composites and paper and pencil factor scores

	ImPACT composites				Paper and pencil factors			
Composite factors	Verbal Memory	Visual Memory	Visual Motor (processing speed)	Reaction Time	Verbal Memory	Visual Memory	Processing Speed	Reaction Time
ImPACT Composites								
Verbal Memory								
Visual Memory	0.358 (.00)							
Visual Motor	0.288 (.04)	0.282 (.04)						
Reaction Time	0.058 (.68)	-0.117 (-.12)	-0.384 (.00)					
Paper and pencil factors								
Verbal Memory	0.400 (.00)	0.189 (.17)	0.107 (.44)	-0.167 (.23)				
Visual Memory	0.438 (.01)	0.590 (.00)	0.213 (.24)	0.138 (.44)	0.383 (.03)			
Processing Spd.	0.117 (.40)	0.183 (.19)	0.409 (.00)	-0.368 (.01)	-0.044 (.75)	-0.137 (.45)		
Reaction Time	-0.103 (.46)	0.220 (.11)	0.342 (.01)	-0.389 (.00)	-0.133 (.34)	0.123 (.50)	0.602 (.00)	

Processing Spd.= processing speed factor score

Table 3. Means and standard deviations (SD) of transformed composite *Z*-scores

Composite/Domain	*N*	Mean	SD
ImPACT Verbal Memory	54	-0.013	1.005
ImPACT Visual Memory	54	-0.029	0.986
ImPACT Visual Motor Speed	54	0.001	1.009
ImPACT Reaction Time	54	0.013	1.005
NP Verbal Memory	54	0.004	0.856
NP Visual Memory	33	0.094	0.770
NP Processing Speed	54	0.164	0.517
NP Reaction Time	54	0.305	0.638

To determine discriminant validity, multi-trait mono-method score combinations were used. First, for each ImPACT and paper-pencil composite score generated, a *z*-score was calculated based on the sample's score distribution (N = 54).[1] Then combinations of the scores were created so that each composite *z*-score could be correlated with the averaged linear combination of the other composites. Reaction time scores were reversed to reflect the fact that a higher score is a worse performance. For instance, the *z*-score for ImPACT Verbal Memory composite was correlated with the average of the *z*-scores for the ImPACT Visual Memory, ImPACT Visual Motor Speed (PS), and ImPACT Reaction Time. The result is a multiply operationalized discriminant validity coefficient for ImPACT Verbal Memory. For example:

$$z1 \ \mathrm{r}((z2 + z3 + z4)/3)$$

where $z1$ is the *z*-score for the Verbal Memory composite, $z2$ is the *z*-score for the Visual Memory composite, $z3$ is the *z*-score for the Visual Motor Speed (processing speed), $z4$ is the *z*-score for the Reaction Time composite, and r is the correlation. The Reaction Time *z*-score was reversed in sign direction.

After computing the multiply-operationalized correlations, the effect of anxiety was considered. In our previous study we found that state anxiety, assessed by self-report during the paper and pencil battery correlated significantly with the ImPACT Visual Motor Speed composite. Thus accounting for that variance was seen as important for this analysis. The NP multi-operationalized correlations were also calculated to serve as a comparison. The means and standard deviations for the composites appear in Table 3.

Results

These data show that, in the current sample, three of the four ImPACT composites were significantly correlated with the other constructs (Visual Memory, Verbal Memory, and Visual Motor Speed), while Reaction Time was not significantly correlated (p = .15: see Table 4). Since the previous study found that state anxiety was significantly correlated with the ImPACT Visual Motor Speed scores (r = –.32, p = .02), it was used as a covariate in the

1. Visual Memory for the paper-pencil battery had fewer participants due to lost data.

Table 4. Point-biserial correlations (*p*-values) of multiply-operationalized ImPACT and paper and pencil battery composite scores (multitrait-monomethod)

Method	Verbal Memory vs Others	Visual Memory vs Others	Visual Motor Speed/Processing Speed vs Others	Reaction Time vs Others
ImPACT: *(*N* = 54)	.275 (.04)	.368 (.01)	.487 (.00)	-.201 (.15)
Paper and pencil: (*N* = 33)	.243 (.17)	.212 (.24)	.200 (.26)	.346 (.05)

*Partial correlations adjusting for state anxiety.

correlations of ImPACT scores. Scores for all cases on the effort measure were acceptable based on the manual for identification of effort (Green Word Memory: Immediate Recall, Delayed Recall, and Consistency were all above 90% for all of the participants).

Although not the focus of the paper, paper and pencil battery results show that all but one of the NP correlations demonstrated discriminant validity as reflected in the lack of statistical significance in the multiply-operationalized score combinations (Reaction Time). The Reaction Time correlation was marginally significant at $p = .05$.

Visual inspection of the correlation matrices confirms that the ImPACT scores are moderately correlated with each other, which was also noted in a previous abstract concerning ImPACT construct validity (Iverson et al., 2003). In the NP sample the high correlation between the processing speed and reaction time scores is notable; the high correlation between visual and verbal memory is obscured in the multiply-operationalized calculation. The differences in sample sizes due to the fewer visual memory scores also likely influence the interpretability somewhat. Finally, visual inspection of the heterotrait–heteromethod correlations (ref. Campbell & Fiske, 1959) shows significant relationships between ImPACT Visual-Motor and NP Reaction Time, and vice versa, as well as ImPACT Verbal Memory and NP Visual Memory. This contradicted the multiple operationalization and suggested that by averaging correlations some (shared trait) variance between specific scores may be obscured.

Discussion

This analysis further supports the findings from Maerlender et al. (2010) in which the specificity of the ImPACT composites was questioned. Using the same sample we demonstrate that only the Reaction Time composite score of ImPACT provided clear indication of construct discriminability, while other composite scores were too highly inter-related to provide meaningful information about specific cognitive functions, despite their labels (e.g., Verbal Memory). Clinically this test would be expected to have adequate sensitivity but limited specificity. One clinical consequence is that clinicians should not make recommendations for treatment based on composite scores alone. For instance, to recommend verbal memory training based on a low composite score would not be appropriate based on these data.

In the NP sample multiply-operationalized discriminability was found for all constructs except Reaction Time. The trend towards significance in the NP Reaction Time score was unexpected and likely reflects the limited task selection and the very high correlation with the Processing Speed score. However, the strong correlation between NP Visual Memory and Impact Verbal Memory raises questions about their discriminability as well.

The use of a multiply-operationalized correlation matrix is atypical and not without its flaws. Psychological constructs reflect ideas and the best one can hope for is an increasing understanding of their structure and properties. A larger sample would be needed to complete a structural equations measurement model, which would provide more robust estimates of score and error variances. Using ImPACT data has its own set of limitations that have been highlighted in many articles, including the validity of individual baseline tests and the self-report of symptoms.

These findings extend and compliment the convergent-construct validity findings from our previous study (Maerlender et al., 2010). That study found that ImPACT composites correlated highly with similar paper and pencil neuropsychological test factors; however, key constructs were diffusely represented (e.g., ImPACT VerbalMemory composite correlated highly with ImPACT and paper and pencil visual memory). Together these studies provide a more complete picture of ImPACT construct validity at baseline which demonstrates strength in its convergent properties, but weakness in the discriminating properties. Thus one might expect construct sensitivity, but limited specificity from this test. We caution against the use of ImPACT alone for the clinical differentiation of specific neuropsychological problems.

Acknowledgments — This work was supported by National Institute of Health BRP R01HD048638 National Center for Medical Rehabilitation Research, National Institute of Child Health and Development, and National Institute of Health RO1NS055020 National Institute of Health/National Institute of Neurological Disease and Stroke. The authors would like to gratefully acknowledge the assistance of Mary Hynes, R.N., Lindley Brainard, Carrie Kruck, M.Psych., Susan Horrigan, M.A., Matthew Garlinghouse, Ph.D., Lindsay Hines, M.A., Laura Hoskins, Psy.D., and John Turco, M.D. The authors would also like to thank the athletic trainers, coaches, and student-athletes at Dartmouth College for their cooperation in this project. A portion of these data was previously presented at the 2008 National Academy of Neuropsychology Annual Meeting in Washington, D.C.

References

Allen, B., & Gfeller, J. (2011). The immediate post-concussion assessment and cognitive testing battery and traditional neuropsychological measures: A construct and concurrent validity study. Brain Injury, 25(2), 179–191.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck D Benedict, R. B. (1997). Brief Visuospatial Memory Test-revised, professional manual. Odessa, FL: Psychological Assessment Resources, Inc.

Campbell, D., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81–105.

Cole, D., Howard, G. S., & Maxwell, S. E. (1981). Effects of mono- versus multipleoperationalization in construct validation efforts. Journal of Counseling and Clinical Psychology, 34, 315–321.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). Delis-Kaplan Executive Function System. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. (2000). California Verbal Learning Test: Manual, 2nd edition. San Antonio, TX: The Psychological Corporation.

Green, P. (2007). The pervasive influence of effort on neuropsychological tests. Physical Medicine and Rehabilitation Clinics of North America, 18, 43–68.

Green, P., & Astner, K. (1995). Manual for the Oral Word Memory Test. Durham, NC: Cognisyst.

Gronwall, D. (1977). Paced auditory serial addition task: A measure of recovery from concussion. Perceptual and Motor Skills, 44, 367–373.

Howard, G., Maerlender, A., Myers, P., & Curtain, T. (1992). In stories we trust: Studies of the validity of autobiographies. Journal of Counseling Psychology, 39, 398–405.

Hunt, T. N., Ferrara, M. S., Miller, L. S., & Macciocchi, S. (2007). Effect of effort on baseline neuropsycholgocial test scores in high school athletes. Archives of Clinical Neuropsychology, 22, 615–621.

Iverson, G. L., Franzen, M. D., Lovell, M. R., & Collins, M. W. (2003). Construct validity of ImPACTTM in athletes with concussion. Presented at the National Academy of Neuropsychology Annual Meeting, Honolulu, HI.

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2005). Validity of ImPACTTM for measuring processing speed following sports-related concussion. Journal of Clinical and Experimental Neuropsychology, 27, 683–689.

Maerlender, A., Flashman, L. A., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., et al. (2010). Examination of the construct validity of ImPACTTM computerized test, traditional, and experimental neuropsychological measures. The Clinical Neuropsychologist, 24, 1309–1325.

McAllister, T. W., Flashman, L. A., McDonald, B. C., & Saykin, A. J. (2006). Mechanisms of working memory dysfunction after mild and moderate TBI: Evidence from functional MRI and neurogenetics. Journal of Neurotrauma, 23, 1450–1467.

Schatz, P., & Putz, B. O. (2006). Cross-validation of measures used for computer based assessment of concussion. Applied Neuropsychology, 13, 151–159.

Smith, A. (1991). Symbol Digit Modalities Test. Los Angeles, CA: Western Psychological Services.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, P. R. (1983). Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, CA: Mind Garden.