Journal of Clinical and Experimental Neuropsychology

ISSN: 1380-3395 (Print) 1744-411X (Online) Journal homepage: http://www.tandfonline.com/loi/ncen20

A pilot investigation of the Motivation Behaviors Checklist (MBC): An observational rating scale of effort towards testing for baseline sports-concussion assessment

Amanda R. Rabinowitz, Victoria Merritt & Peter A. Arnett

To cite this article: Amanda R. Rabinowitz, Victoria Merritt & Peter A. Arnett (2016) A pilot investigation of the Motivation Behaviors Checklist (MBC): An observational rating scale of effort towards testing for baseline sports-concussion assessment, Journal of Clinical and Experimental Neuropsychology, 38:6, 599-610, DOI: 10.1080/13803395.2015.1123224

To link to this article: http://dx.doi.org/10.1080/13803395.2015.1123224

Published online: 10 Apr 2016.

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY, 2016
VOL. 38, NO. 6, 599–610
http://dx.doi.org/10.1080/13803395.2015.1123224

Routledge
Taylor & Francis Group

A pilot investigation of the Motivation Behaviors Checklist (MBC): An observational rating scale of effort towards testing for baseline sports-concussion assessment

Amanda R. Rabinowitz[a], Victoria Merritt[b] and Peter A. Arnett[b]

[a]Moss Rehabilitation Research Institute, Elkins Park, PA, USA; [b]Psychology Department, The Pennsylvania State University, University Park, PA, USA

ABSTRACT

Background: Baseline neuropsychological testing is commonly used in the management of sports-related concussion. However, underperformance due to poor effort could lead to invalid conclusions regarding postconcussion cognitive decline. We designed the Motivation Behaviors Checklist (MBC) as an observational rating scale to assess effort towards baseline neuropsychological testing. Here we present preliminary data in support of its reliability and validity. *Method*: MBC items were generated based on the consensus of a panel of graduate students, undergraduates, and a clinical neuropsychologist who conduct neuropsychological evaluations for a sports concussion management program. A total of 261 college athletes were administered a standard neuropsychological test battery in addition to the MBC. A subset of evaluations (n= 101) was videotape and viewed by a second rater. Exploratory factor analysis (EFA) was used to refine the scale, and reliability and validity were evaluated. *Results*: EFA revealed that the MBC items represent four latent factors—Complaints, Poor Focus, Psychomotor Agitation, and Impulsivity. Reliability analyses demonstrated that the MBC has good inter-rater reliability (intraclass correlation coefficient, ICC = .767) and internal consistency (α = .839). The construct validity of the MBC is supported by large correlations with examiners' ratings of effort (ρ = –.623) and medium-sized relationships with cognitive performance and self-ratings of effort ($|\rho|$ between .263 and .345). Discriminant validity was supported by nonsignificant correlations with measures of depression and postconcussion symptoms (ρ = .056 and .082, respectively). *Conclusions*: These findings provide preliminary evidence that the MBC could be a useful adjunct to baseline neuropsychological evaluations for sports-concussion management.

ARTICLE HISTORY
Received 12 May 2015
Accepted 18 November 2015

KEYWORDS
Concussion; effort; mild traumatic brain injury; motivation; baseline testing; performance validity

Baseline neuropsychological testing for the management of sports-related concussion has gained widespread popularity. Although the availability of preinjury cognitive data is intuitively appealing, consensus statements have cautioned that there is insufficient evidence to support the use of routine baseline testing (McCrory et al., 2013). Others have gone further, arguing that baseline testing does nothing to reduce risk of reinjury and may increase risk of premature return to play in some cases (Kirkwood, Randolph, & Yeates, 2009; Randolph, 2011).

A major pitfall of baseline concussion testing is athletes' effort towards testing. Although athletes are highly motivated to put forth their best effort after a concussion, they have little incentive to perform well when baseline tests are administered at the beginning of the athletic season. In fact, there can be an incentive to underperform at baseline, as savvy athletes recognize that their baseline test data will be used to determine return to play should they sustain a concussion. This phenomenon, called sandbagging, refers to a deliberate attempt to misrepresent one's preinjury cognitive status. Studies in which test-takers are instructed to fake poor performance on computerized cognitive tests have shown that embedded performance validity indicators capture the majority of intentional underperformers (Erdal, 2012; Schatz & Glatts, 2013).

CONTACT Amanda R. Rabinowitz rabinowitz.a@gmail.com Moss Rehabilitation Research Institute, 50 Township Line Road, Elkins Park, PA 19027, USA.

Sandbagging is a purposeful mechanism of cognitive underperformance. However, even athletes who approach testing in an honest and straightforward manner may exhibit suboptimal performance when they are not properly motivated to put forth their best effort. Hence, cognitive underperformance could also be influenced by passive unintentional mechanisms (such as boredom, distractions within the testing environment, or fatigue from a recent workout or practice). An athlete's adequate, but suboptimal, baseline performance could serve as a misleading benchmark by which her highly motivated postconcussion performance is judged.

Research has demonstrated that baseline performance can be influenced by poor effort (Bailey, Echemendia, & Arnett, 2006; Hunt, Ferrara, Miller, & Macciocchi, 2007; Rabinowitz, Merritt, & Arnett, 2015; Schatz, 2010; Solomon & Haase, 2008). Furthermore, emerging evidence from the field of positive psychology suggests that individual differences in motivation influence cognitive performance in other contexts. Duckworth and colleagues conducted a meta-analysis of randomized experiments testing the effects of material incentives on intelligence-test performance and found that incentives increased IQ scores by an average of 0.64 standard deviations (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011). In a follow-up study, the researchers trained observers to rate the motivation levels of 251 adolescent boys by viewing 15-min "thin slice" videos of IQ test administrations and found that observer ratings of test motivation were associated with test performance (Duckworth et al., 2011).

Suspect effort at baseline poses a critical problem for the use of neuropsychological assessment in concussion management. Assessors are at risk of underestimating the cognitive effects of a concussion if estimates of athletes' preinjury aptitude are invalid. Hence, there is a need for measures of effort towards testing that can provide context to aid in the interpretation of pre- and postinjury neuropsychological test results. Although there are a number of measures of performance validity—the act of "faking" impaired performance on a neuropsychological evaluation—passive mechanisms of underperformance, as discussed above, have yet to be well defined and may go undetected by standard performance validity tests.

The Immediate Post-Concussion Assessment and Cognitive Test (ImPACT; Lovell, Collins, Podell, Powell, & Maroon, 2000) is a popular computerized baseline and postconcussion testing battery, currently used to help manage concussions in professional, collegiate, and high-school athletes. In order to insure the integrity of baseline assessments, the ImPACT test creators have established "red flags" and validity indicators for poor performance, generally based on scores 2 standard deviations (*SD*s) below the mean on certain test indices. One study examined the sensitivity of these validity indicators by instructing athletes to deliberately underperform and found that some athletes were able to fake lower scores without detection (Erdal, 2012). These findings suggest that performance-based validity indicators may fail to capture all cases of faking cognitive deficits, which, in theory, should have a more pronounced influence on test performance than poor effort without the intent to deceive.

Furthermore, it is possible that incremental changes in effort level may have an important influence on neuropsychological test performance in a way that goes unappreciated by categorical approaches to effort testing that flag patients' profiles as "suspect" or "invalid." For example, imagine a college freshman joining the women's soccer team with a true premorbid verbal memory ability of 120 standard score points. She comes to the baseline testing session directly after a rigorous workout. The purpose of baseline testing has been explained to her; however, some details have been forgotten amidst the demands of acclimating to her new campus, team, and schedule of courses. She is surprised to learn that the testing session will take an hour, which interferes with her plans to purchase textbooks before meeting teammates for dinner. The examiner encourages her to put forth her best effort. She is friendly and cooperative, although slightly fatigued and distracted. On a test of verbal memory she receives a score of 100 standard score points, underestimating her true premorbid ability of 120 standard score points. Overall, her baseline test performance is well within the normal range and does not trigger performance invalidity indicators.

Later that season, she sustains a concussion and experiences problems with memory. Despite her deficit, she is highly motivated to perform well on the postconcussion assessment. She is eager to return to play and wants to present an accurate impression of her current cognitive functioning so that her team doctor can make an informed return-to-play decision. She receives a verbal memory score of 100 standard score points, technically "back to baseline," but still substantially below her true premorbid ability level. This hypothetical scenario demonstrates that valid cognitive performance may not be optimal. Changes in approach to testing may be important to consider, as discrepancies between optimal performance and performance that is merely "valid" could have implications for postconcussion management. Continuous measures are necessary to capture incremental changes in effort and detect discrepancies between baseline and postconcussion assessments.

With these considerations in mind, the aim of this study was to develop a scale that can be used as a measure of athletes' effort towards baseline neuropsychological testing. Our goal was to create an observational rating scale meeting the following specifications: (a) It consists of objective behaviors that indicate effort towards testing, (b) it is sensitive to a range of effort levels from poor to exceptional effort, and (c) it can be administered quickly and easily with minimal training. Based on these criteria, we generated items and administered the resultant scale (referred to subsequently as the Motivation Behaviors Checklist: MBC) as part of the sports concussion management program at a Division 1 university. Here we present preliminary evidence for the reliability and validity of the MBC within the context of individualized baseline neuropsychological evaluation of college athletes.

Table 1. Complete list of 26 MBC items.

Item no.	Item content
1	Demonstrates error-related adjustments in approach to testing.
2	Responds impulsively or carelessly.
3	Is nonresponsive (blank expression).
4	Scrunches facial muscles (indicating concentration).
5	Orients body or test materials as to "tune out" extraneous stimuli.
6	Uses motor behaviors to assist with test performance (counts, traces or taps with fingers, looks or tilts head up).
7	Asks for instructions to be repeated.
8	Makes good eye-contact.
9	Asks clarification questions.
10	Eyes wander around the room.
11	Uses cell-phone, or looks at cell phone during testing.
12	Groans or sighs.
13	Slumps down in chair.
14	Puts head down on the table.
15	Engages the examiner in test-irrelevant conversation during testing.
16	Demonstrates interest in how well s/he is performing.
17	Asks questions about how long testing is, "are we almost done."
18	Mentions being bored or distracted.
19	Fidgets with pencils or other materials.
20	Has sloppy and careless hand-writing when writing or drawing.
21	Appears to take an adequate amount of time to complete tasks.
22	Follows directions.
23	Rolls eyes.
24	Expresses that they "don't want to be here."
25	Complains of being tired or hungry.
26	Moves in more closely when a stimulus is visually or verbally presented.

Note. MBC = Motivation Behaviors Checklist. Items are rated on a 5-point Likert scale where 1 indicates that the behavior was "never" exhibited, and 5 indicates that the behavior was "frequently" exhibited.

Method

Scale development

Items were generated based on the consensus of a panel composed of a clinical neuropsychologist and licensed psychologist (P.A.) and 13 graduate and undergraduate research assistants directly involved in baseline and postconcussion neuropsychological assessment of college athletes. In discussions of motivation behaviors, the panel was instructed to think of observable behaviors that indicate both high and low levels of effort towards testing. As a result of this endeavor, 26 behaviors were identified. From these behaviors, we created 26 Likert-scale items that measure the frequency of each behavior, ranging from "never exhibits the behavior" to "frequently exhibits the behavior." (See Table 1 for the complete list of items.)

Participants

Participants in the present study were 261 college athletes participating in a multisport concussion management program at a large state university, and 40 uninjured control participants who played intramural sports but did not participate in the concussion management program. This program is designed according to the Sports as a Laboratory Assessment Model (SLAM; Barth, Harvey, Freeman, Brosheck, 2011) paradigm, in order to provide objective neuropsychological test-data to team physicians to inform return-to-play decisions. Eight athletic teams regularly participate in baseline testing: football, men's and women's lacrosse, men's and women's soccer, men's and women's basketball, men's ice hockey, and wrestling. Baseline and postconcussion testing sessions are administered in a one-on-one format, with paper-and-pencil tests and computerized testing taking place within the same session. Postconcussion testing is provided for any athlete who sustains a head injury. Whenever possible, postconcussion assessment takes place within 48 hours of injury.

Validity and internal consistency analyses were conducted in a sample of 261 athletes. Only baseline data were included in these analyses (baseline MBC sample). A subset of evaluations (n = 101) were videotaped and were viewed by a second rater who was blind to the initial MBC and effort ratings. This sample was composed of 40 control participants, 52 athletes at baseline, and 9 athletes post concussion (reliability sample). Raters were drawn from the same pool of graduate and undergraduate assistants as those who administered baseline and postconcussion neuropsychological evaluations. Raters were randomly assigned to cases, with the only exclusion rule being that the second rater could not have conducted the athlete's baseline testing session (i.e., could not have been the original MBC rater). Reliability ratings were equally distributed across the pool of raters. The second rater viewed the entire videotaped testing session and then completed the 26-item MBC and the examiner's rating of effort based on their observations of the videotaped testing session.

Measures

The test battery consists of a number of measures that assess cognitive, physiological, and affective functioning, including: the Hopkins Verbal Learning Test–Revised (HVLT–R; Benedict, Schretlen, Groninger, & Brandt, 1998), the Brief Visuospatial Memory Test–Revised (BVMT–R; Benedict, 1997), the Symbol–Digit Modalities Test (SDMT; Smith, 1982), the Digit Span Test (Wechsler, 1997), the PSU Cancellation Task, the Stroop Color–Word Test (SCWT; Trenerry, Crosson, DeBoe, & Leber, 1989), Controlled Oral Word Association (COWA; Ruff, Light, Parker, & Levin, 1996), Comprehensive Trailmaking Test (CTMT; Reynolds, 2002), the Wechsler Test of Adult Reading (WTAR; Psychological Corporation, 2001), and the BDI-Fast Screen (BDI-FS; Beck, Steer, & Brown, 2000). The Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997), a forced-choice digit recognition test, was administered as a performance validity test. We used a cutoff of 89% or less as a marker of inadequate effort (Conder & Allen, 1999). The ImPACT (Lovell et al., 2000), was also used. The ImPACT is a computerized test battery consisting of three main parts: demographic data, neuropsychological tests, and the Post-Concussion Symptom Scale. Six neuropsychological tests are included, designed to target memory, processing speed, and reaction time. From the six tests, five composite scores are derived: verbal memory, visual memory, visuomotor speed, reaction time, and impulse control. The sum of all items from the ImPACT Post-Concussion Symptom Scale forms the ImPACT Total Symptom Composite.

Examiner ratings

In addition to completing the MBC, examiners were asked to complete a single Likert-scale item describing the level of effort exhibited by the athlete during testing, where a rating of 1 indicated "not trying at all," and a rating of 7 indicated "optimal effort" (examiner's rating of effort). Athletes also rated their own level of effort towards testing on the same scale (self-rating of effort). Examiners were instructed to assign ratings based on how hard the participant was trying, and not how well they performed on tests. Examiners also completed a single-item affect rating. They were instructed to rate each participant on a scale of 1 (dysphoria) to 5 (elation) according to manifest affect throughout the evaluation.

Procedure

Athletes completed baseline neuropsychological testing, including the measures described above. Evaluations were conducted by a PhD-level clinical neuropsychologist, or a graduate or undergraduate assistant who had been trained by a PhD-level clinical neuropsychologist. Testing sessions took approximately 1.5 hours per evaluation and also required approximately 30 min for paperwork, debriefing, and administration of other instruments not included in the present study. At the end of the testing session, test administrators completed the 26-item MBC and other examiner ratings based on their observations during testing. Written informed consent was obtained from study participants. The study received approval from The Pennsylvania State University

Institutional Review Board and was conducted in accordance with its ethical guidelines.

Approach to data analysis

Reliability and validity analyses were conducted using an iterative process. First inter-rater reliability and internal consistency were evaluated. Then exploratory factor analysis (EFA) was conducted to decompose the scale into its latent constructs and remove items that loaded poorly on any factor. All EFA procedures were developed a priori. Factors were extracted by applying the criterion, eigenvalue > 1 (Kaiser, 1958). We rotated the factors using an oblique rotation (direct oblimin with Kaiser normalization). We chose an oblique rotation because we had no a priori hypothesis that the MBC factors would be unrelated but, rather, assumed that they would be correlated, as most psychological constructs are (Floyd & Widaman, 1995; Schmitt, 2011; Worthington & Whittaker, 2006). Items were retained if they exhibited factor loadings >.4 on one of the factors (Worthington & Whittaker, 2006). Because we utilized an oblique factor rotation, which allows factors to correlate with one another, we did not select items for deletion based on cross-loadings (Schmitt, 2011). Regression-based factor scores were calculated using SPSS v. 21. The factor score is the sum of standardized observed item values that have been weighted by regression coefficients. Regression coefficients are obtained by multiplying the inverse of the observed variable correlation matrix by the matrix of factor loadings and the factor correlation matrix (DiStefano, Zhu, & Mindrila, 2009).

Next, the construct validity of each latent factor was assessed by examining its correlations with a priori validity criteria: a subset of cognitive performance variables and the examiner and self-ratings of effort. Factors that poorly predicted the criterion variables were discarded, and a revised MBC scale was calculated composed of the retained items. The process described above was then repeated using the revised MBC—including reliability analyses, EFA, and validity analyses. All statistics were conducted using SPSS v. 21.

The selection of appropriate validity criteria posed a challenge, as there are currently no validated measures of effort as it is defined in the current investigation. The single-item examiner and self-ratings of effort were developed to serve as face-valid measures of effort. Based on the assumption that effort should influence cognitive performance, we wanted to evaluate cognitive outcome measures as validity criterion variables as well. A number of cognitive outcome variables could be derived from the test battery administered in the current study. We selected the ImPACT composite scores as validity criteria for several reasons. First, the test battery contains considerable redundancy, with multiple measures of memory, attention, and processing speed. We wanted to limit the total number of statistical comparisons in our validity analysis by selecting only one variable for each cognitive construct. Second, because the ImPACT is administered and scored via computer, the examiner did not have access to the test-taker's performance at the time that the MBC ratings were completed. This is in contrast to paper-and-pencil tests, for which the examiner could reasonably gauge how well the test-taker performed during test administration. Hence, in theory, there should be less cross contamination between ImPACT test performance and MBC ratings than between paper-and-pencil test performance and MBC ratings. Third, our intention was to create a scale that would be useful in sports concussion management —specifically, for the purpose of interpreting baseline cognitive test performance. The ImPACT is commonly used in the clinical management of sports-related concussion, particularly in clinics and programs that make use of the baseline-testing model. Hence, findings related to ImPACT performance may be relevant to many practitioners who utilize baseline test data.

We conceptualize effort as a related but distinct phenomenon from performance validity. However, to provide context, we compared MBC scores for athletes who passed and failed existing performance validity measures—the CARB, the ImPACT "red flags," and the ImPACT validity indicators.

Results

Demographic characteristics of the sample are available in Table 2. Items 1, 4, 5, 6, 7, 8, 9, 16, 21, 22, and 26 were reverse coded, and an MBC summary score was then calculated by summing the 26 items. Higher scores reflect poorer effort. Inter-rater reliability analysis was conducted on a subsample of athletes with videotaped test sessions (reliability sample, n = 101). Inter-rater reliability of the single-item examiners' rating of effort was

Table 2. Demographic characteristics of the sample.

	Baseline MBC sample (*N* = 261)				Reliability sample (*N* = 101)			
Variables	*M*	*SD*	*N*	%	*M*	*SD*	*N*	%
Age (years)	18.74	1.33			18.90	1.27		
Sex								
Male			196	75.1			66	65.3
Female			65	24.9			35	34.7
Ethnicity								
Caucasian			192	73.6			75	74.3
African American			52	19.9			8	7.9
Other			17	6.5			18	17.8
Sport								
Football			51	19.5			1	1.0
Men's basketball			20	7.7			6	5.9
Men's ice hockey			31	11.9			1	1.0
Men's lacrosse			51	19.5			17	16.8
Men's soccer			29	11.1			10	9.9
Women's basketball			14	5.4			4	4.0
Women's ice hockey			3	1.1			0	0
Women's lacrosse			8	3.1			1	1.0
Women's soccer			38	14.6			7	6.9
Wrestling			15	5.7			14	13.9
Other sport			1	0.4			3	3.0
No sport			0	0			37	36.6

Note. MBC = Motivation Behaviors Checklist.

assessed via single measure intraclass correlation (ICC) using a one-way random effects model, which revealed adequate agreement between raters: ICC = .631, 95% confidence interval (CI) = [.476, .748], $F(76, 77) = 4.418$, $p < .001$. Inter-rater reliability of the 26-item MBC was assessed via scale-level ICC using a one-way random effects model, which revealed good agreement between raters: ICC = .779, 95% CI = [.669, .852], $F(95, 96) = 4.524$, $p < .001$. Internal consistency was assessed in the MBC baseline sample ($n = 261$) using Cronbach's α, which revealed adequate internal consistency of the full 26-item MBC (α = .606).

Next, the 26 items were subject to factor analysis using maximum likelihood estimation. Direct oblimin rotation with Kaiser normalization was utilized. The Kaiser–Meier–Olkin Test of Sampling Adequacy (KMO) suggested that the shared variance among items was meritorious for factor analysis (KMO = .824). Eight factors with eigenvalues greater than 1 were extracted, explaining 64% of the variance among items. Items with factor loadings less than .40 were removed from the scale. Based on this criterion, four items were removed from the item pool: "Puts head down on the table," "Is nonresponsive (blank expression)," "Uses cell phone, or looks at cell phone during testing," "Demonstrates error-related adjustments in approach to testing." Inter-rater reliability of the 22-item MBC was assessed via scale-level ICC using a one-way random effects model, which revealed good agreement between raters: ICC = .766, 95% CI = [.651, .844], $F(95, 96) = 4.524$, $p < .001$. Internal consistency was assessed in the MBC baseline sample ($n = 261$) using Cronbach's α, which revealed good internal consistency of the 22-item MBC (α = .796). The 22 items were then subject to a second factor analysis using maximum likelihood estimation, and factors were rotated using a direct oblimin rotation with Kaiser normalization. The KMO test suggested that the shared variance among items was meritorious for factor analysis (KMO = .823). Seven factors with eigenvalues greater than 1 were extracted, explaining 69% of the variance among items. All remaining items had factor loadings greater than .40 (see Table 3 for items and their factor loadings).

Table 3. Factor loadings for 22-item MBC.

	Factor						
MBC item	1	2	3	4	5	6	7
Moves in more closely (r)	.998						
Good eye contact (r)	.990						
Don't want to be here		.851					.448
Tired hungry		.791					.484
Mentions bored distracted		.720				.410	
Asks about length of testing		.592		−.418		.570	
Uses motor behavior to assist performance (r)			.749				
Scrunches facial muscles (r)			.717				
Tunes out extraneous stimuli (r)			.659				
Test-irrelevant conversation				−.669			
Interest in performance (r)				.668			
Asks for repeated instructions (r)					.841		
Asks clarification questions (r)					.715		
Slumps down						.720	
Groans sighs		.524				.711	
Eyes wander						.622	
Rolls eyes						.450	
Fidgets						.429	.407
Follow directions (r)		.433					.655
Takes adequate time (r)							.603
Responds impulsively		.478				.418	.568
Sloppy handwriting		.438					.515

Note. MBC = Motivation Behaviors Checklist. Items marked (r) are reverse-coded.

The items were summed to create a total score for the 22-item scale. Nonparametric correlations between the 22-item MBC and validity criteria were calculated to determine the convergent and discriminant validity. The 22-item MBC scale demonstrated statistically significant correlations with all validity criteria variables, including ImPACT verbal memory composite, ImPACT visual memory composite, ImPACT visual motor speed composite, ImPACT reaction time composite, self-rating of effort, and examiner's rating of effort (see Table 4). In order to establish discriminant validity, correlations were examined between the MBC scale and three variables presumed to be unrelated to effort—depression (BDI-FS), postconcussion symptoms (Postconcussion Symptom Scale, PCSS), and the single-item examiner affect rating. MBC score was not significantly related to BDI-FS or PCSS, but was significantly correlated with examiner affect ratings, such that greater negative affect was associated with poorer effort towards testing (see Table 4 for complete correlation results).

In attempts to simplify the scale and improve validity, we examined the validity of each of the 22-item MBC 7 factors. Regression-based factor scores for each of the seven factors were calculated for each subject, and their relationships to validity criteria were examined using nonparametric correlation (Table 5). Based on this analysis, two of the seven factors (Factor 4 and Factor 5) were deemed poor predictors of the validity criterion. The following four items from Factors 4 and 5 were removed from the item pool: "Engages the examiner in test-irrelevant conversation," "Demonstrates interest in how well s/he is performing," "Asks for instructions to be repeated," and "Asks clarification questions." Cronbach's alpha revealed good internal consistency of the remaining 18 items (α = .839), and inter-rater reliability of the 18-point scale was good: ICC = .767, 95% CI [.652, .844], $F(96, 97) = 4.287$, $p < .001$. The surviving 18 items were subject to factor analysis using maximum likelihood estimation, and factors were rotated using a direct oblimin rotation with Kaiser normalization. The KMO test suggested that the shared variance among items was meritorious for factor analysis (KMO = .846). Four factors with eigenvalues greater than 1 were extracted explaining 59% of the variance among items. All remaining items had factor loadings greater than .40 (see Table 6 for items and their factor loadings). Visual inspection of the items suggested that the factors represent the following latent constructs: (1) Complaints; (2) Poor Focus; (3) Psychomotor Agitation; and (4) Impulsivity.

The final 18-item MBC scale demonstrated statistically significant correlations with all validity criteria variables, including ImPACT verbal memory composite (ρ = −.342, $p < .001$, medium effect size), ImPACT visual memory composite (ρ = −.304, $p < .001$, medium effect size), ImPACT visual motor speed composite (ρ = −.263, $p < .001$, medium effect

Table 4. Nonparametric correlations of the MBC total scale scores with validity criteria.

	MBC		
Validity criterion	26-Item	22-Item	18-Item
Convergent validity			
Verbal memory	−.293**	−.291**	−.342**
Visual memory	−.253**	−.277**	−.304**
VM speed	−.239**	−.227**	−.263**
Reaction time	.301**	.302**	.345**
Self-motivation	−.247**	−.242**	−.269**
Examiner effort	−.620**	−.615**	−.623**
Discriminant validity			
BDI-FS	.040	.046	.056
PCSS	.093	.071	.082
Examiner affect	−.484**	−.457**	−.451**

Note. Correlations: Spearman's rho. MBC (Motivation Behaviors Checklist) total scale scores: 26-, 22-, and 18-item versions. Verbal memory = ImPACT (Immediate Post-Concussion Assessment and Cognitive Test) verbal memory composite; visual memory = ImPACT visual memory composite; VM speed = ImPACT visual memory composite; reaction time = ImPACT reaction time composite; self-motivation = self-motivation rating; examiner effort = examiner's rating of effort; BDI-FS = Beck Depression Inventory–Fast Screen; PCSS = ImPACT Total Symptom Scale; examiner affect = examiner affect rating.
**$p < .01$.

Table 5. Nonparametric correlations of 7 MBC factors and validity criteria.

Validity criterion	Factors						
	1	2	3	4	5	6	7
Verbal memory	−.27**	−.19**	−.27**	−.05	.15*	−.20**	−.28**
Visual memory	−.25**	−.06	−.25**	−.06	.06	−.16**	−.23**
VM speed	−.18**	−.17**	−.11	−.03	.13*	−.25**	−.22**
Reaction time	.23**	.20**	.18**	−.03	−.13*	.26**	.25**
Self-effort	−.16**	−.21**	−.09	.01	.05	−.24**	−.11
Examiner effort	−.47**	−.37**	−.46**	−.11	.01	−.46**	−.36**

Note. Correlations: Spearman's rho. MBC = Motivation Behaviors Checklist; verbal memory = ImPACT (Immediate Post-Concussion Assessment and Cognitive Test) verbal memory composite; visual memory = ImPACT visual memory composite; VM speed = ImPACT visual memory composite; reaction time = ImPACT reaction time composite; self-effort = self-rating of effort; examiner effort = examiner's rating of effort.
*$p < .05$. **$p < .01$.

Table 6. Factor loadings for 18-item MBC.

	Factor			
MBC item	1	2	3	4
Don't want to be here	.806			.533
Tired hungry	.771			.542
Mentions bored distracted	.739			
Asks about length of testing	.628		−.581	
Uses motor behavior to assist performance		.750		
Tunes out extraneous stimuli		.687		
Scrunches facial muscles		.684		
Moves in more closely		.603		
Slumps down			−.730	
Groans sighs	.516		−.696	
Eyes wander			−.612	
Rolls eyes			−.454	
Follow directions				.717
Responds impulsively	.426			.607
Takes adequate time				.555
Sloppy handwriting				.535
Good eye contact				.498
Fidgets				.451

Note. MBC = Motivation Behaviors Checklist.

size), ImPACT reaction time composite (ρ = .345, $p < .001$, medium effect size), self-rating of effort (ρ = −.269, $p < .001$, medium effect size), and examiner's rating of effort (ρ = −.623, $p < .001$, large effect size). In order to establish discriminant validity, correlations were examined between the MBC scale and three variables presumed to be unrelated to effort—depression (BDI-FS), postconcussion symptoms (PCSS), and the single-item examiner affect rating. MBC score was not significantly related to BDI-FS (ρ = .056, $p = .375$, small effect size) or PCSS (ρ = .082, $p = .191$, small effect size). MBC was significantly correlated with examiner affect ratings (ρ = −.451, $p < .001$, large effect size), such that greater negative affect was associated with poorer effort towards testing. A comparison of reliability, factor structure, and validity criteria for 26-item, 22-item, and 18-item versions of the MBC is presented in Table 7. Complete validity findings are presented in Table 4. Both the 22-item 7-factor and 18-item 4-factor versions of the MBC exhibited good reliability and validity. The 18-item MBC is favored because of its relatively shorter length and simpler structure. (For the 18-item MBC see Table 8.)

Descriptive statistics were examined for the final 18-item MBC scale—in the MBC baseline sample ($n = 261$) the MBC has a mean (*SD*) of 34.63 (9.04) and a median 33. The possible range of scores is 18 to 90. The observed minimum in our sample was 18, and the observed maximum was 70. (See Figure 1 for a histogram of MBC scores.) The distribution of scores is non-normal according to the Kolmogorov–Smirnov test (0.124, $p < .001$) and exhibited a left-skewed (skewness = 1.086, *SE* = 0.151) and platykurtic (kurtosis = 1.711, *SE* = 0.301) distribution.

We compared MBC scores of athletes who triggered one of the ImPACT's "red flags" or validity indicators with those who did not. Fifty-six of the 261 athletes exhibited one of the ImPACT's four "red flags"—(a) processing speed composite < 25, (b) reaction time composite < 0.8 s, (c) verbal memory composite < 70%, and (d) visual memory composite < 60%. Athletes with one or more "red flags" had MBC scores that were 4.66 points higher (thus indicating more behaviors indicative of low motivation), on average, than those without "red flags," $t(256) = -3.484$, $p = .001$. Only nine of the 261 athletes failed one of the ImPACT's five validity indicators—(a) Xs and Os total interference incorrect < 30, (b) impulse control composite < 30, (c) word memory learning percentage correct < 69%, (d) design memory learning percentage correct < 50%, and (e) three letters total letters correct < 8. Athletes who failed one or more validity indicators had MBC scores that were 7.43 points higher, on average, than those who passed all five validity indicators, $t(255) = -2.460$, $p = .015$. Only 14 of the 261 athletes failed the CARB (with scores of 89% or less correct). There was no statistically significant difference in MBC score between

Table 7. Comparative reliability, validity, and factor structure for the 26-, 22-, and 18-item versions of the MBC.

	Reliability analysis		Factor structure		Validity analysis	
Items (*N*)	Cronbach's α	Inter-rater	Number of factors	Variance explained %	Convergent validity	Discriminant validity
26	.606	.779	8	64	.24 to .62	.04 to .48
22	.796	.766	7	69	.23 to .62	.05 to .46
18	.839	.767	4	59	.26 to .62	.06 to .45

Note. MBC = Motivation Behaviors Checklist; inter-rater = scale-level intraclass correlation; convergent validity = the range of the absolute value of Spearman's rho for correlations between MBC and convergent validity variables (ImPACT verbal memory composite, ImPACT visual memory composite, ImPACT visual memory composite, ImPACT reaction time composite, self-motivation rating, and examiner motivation rating; ImPACT = Immediate Post-Concussion Assessment and Cognitive Test); discriminant validity = the range of the absolute value of Spearman's rho for correlations between MBC and discriminant validity variables (Beck Depression Inventory–Fast Screen, ImPACT Total Symptom Scale, and examiner affect rating).

Table 8. Final version of the MBC.

Item number		
MBC-18	MBC-26	Item content
1	2	Responds impulsively or carelessly.
2	4	Scrunches facial muscles (indicating concentration). (r)
3	5	Orients body or test materials as to "tune out" extraneous stimuli. (r)
4	6	Uses motor behaviors to assist with test performance (counts, traces or taps with fingers, looks or tilts head up). (r)
5	8	Makes good eye-contact. (r)
6	10	Eyes wander around the room.
7	12	Groans or sighs.
8	13	Slumps down in chair.
9	17	Asks questions about how long testing is, "are we almost done."
10	18	Mentions being bored or distracted.
11	19	Fidgets with pencils or other materials.
12	20	Has sloppy and careless hand-writing when writing or drawing.
13	21	Appears to take an adequate amount of time to complete tasks. (r)
14	22	Follows directions. (r)
15	23	Rolls eyes.
16	24	Expresses that they "don't want to be here."
17	25	Complains of being tired or hungry.
18	26	Moves in more closely when a stimulus is visually or verbally presented. (r)

Note. MBC = Motivation Behaviors Checklist. Raters assign each Likert item a rating from 1–5, with 1 indicating that the athlete "never exhibits the behavior" and 5 indicating that the athlete "frequently exhibits the behavior." MBC-18 = final 18-item MBC Scale; MBC-26 = original 26-item MBC Scale. Items marked (r) are reverse-coded.

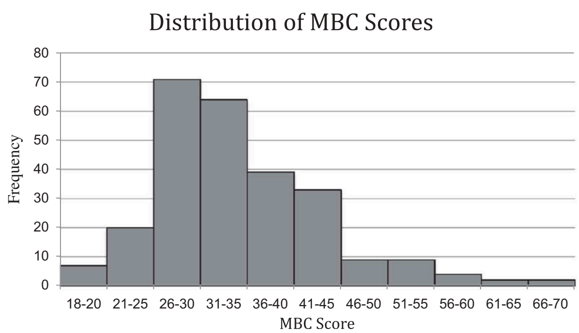

Figure 1. Histogram of Motivation Behaviors Checklist (MBC) scores.

athletes who passed (*M*= 34.47, *SD*= 9.12) and those who failed (*M*= 36.86, *SD*= 7.71) the CARB, $t(250) = -0.957$, $p = .339$.

As a post hoc analysis, we examined a score of 53 on the MBC (2 *SD*s above the aggregate mean of the sample of athletes at baseline) as a possible cutoff score indicating suspect baseline performance. Fourteen of the 261 athletes had MBC scores greater than or equal to 53. Of these 14 athletes, five had CARB scores less than 100%, five triggered one or more of the ImPACT's "red flags," and two triggered one or more validity indicators. In all, seven of the 14 athletes with suspect MBC scores also exhibited at least one of the other markers of poor effort listed above.

Discussion

We designed the MBC—an observational checklist that assesses effort towards testing—to be used in baseline neuropsychological evaluation for sports-concussion management. Here we present preliminary data in support of its reliability and validity. Factor analysis revealed that the retained MBC items represent four latent factors—Complaints, Poor Focus, Psychomotor Agitation, and Impulsivity. Reliability analyses demonstrated that the 18-item MBC has good inter-rater reliability and good internal consistency. The construct validity of the MBC is supported by a correlation with examiners' ratings of effort and medium-sized relationships with cognitive performance and self-ratings of effort. Discriminant validity was supported by a lack of correlation with measures of depression and postconcussion symptoms. Of note, there was a significant correlation between examiners' single-item ratings of the athletes' affect and MBC scores, suggesting convergent validity between measures. Shared method variance is a likely contributor to the relationship between these variables.

We compared MBC scores for athletes who passed and failed existing performance validity measures—the CARB and the ImPACT's embedded performance validity indicators and "red flags." We found that athletes who triggered the ImPACT's "red flags" and validity indicators had higher MBC scores, suggesting poor effort towards testing. Although there was a small numerical difference in the expected direction, MBC scores for athletes who failed the CARB did not significantly differ from those of athletes who passed. As might be expected, these results suggest that there is overlap between effort towards testing, as measured by the MBC, and performance validity, as measured by performance-based effort tests. However, the lack of association between the MBC and the CARB indicates that response bias, as measured by the CARB, may be distinct from cognitive underperformance due to passive unintentional mechanisms, such as boredom, fatigue, and so on. This is not entirely surprising, because, as hypothesized, suboptimal effort may be conceptually distinct from intentional underperformance.

These preliminary findings suggest that the MBC could be a useful adjunct to baseline neuropsychological evaluations for sports-concussion management. With further validation to establish cutoff criteria for baseline MBC and pre–postinjury MBC change scores, we believe that the MBC could provide important information that would aid in the interpretation of baseline neuropsychological test results. Our results are consistent with those of Duckworth and colleagues (2011), who also found that examiner ratings of effort predicted cognitive performance. In the present sample we found that MBC ratings explained roughly 9% of the variance in each of four cognitive performance variables examined. This finding adds to the body of evidence suggesting that incentives and nonintellective traits have a meaningful influence on cognitive performance.

The most serious limitation of the present study is the challenge of selecting appropriate criteria for evaluating the validity of our scale. In creating this scale, we hoped to develop a measure of effort that captures subtle forms of cognitive underperformance due to poor motivation. We argue that this is a distinct phenomenon from cognitive underperformance with intent to deceive. In lieu of a "gold standard" effort measure to use as a validity criterion, we took a broad approach by examining the MBC in relationship to cognitive performance, observation ratings, self-ratings, and embedded and external performance validity indicators.

The 18-item observational scale and examiners' single-item effort ratings are conceptually distinct —the MBC items were designed to be objective (examiners rate the frequencies of different observable behaviors) whereas the single item motivation rating is a general impression of the athletes' effort level. However, there is an inherent circularity in comparing two observational ratings from the same rater. The effort self-rating does not have this same problem with circularity, but it is limited by the introspective capacity of the test-taker. We selected the ImPACT composite scores as our cognitive performance measures to minimize the influence of performance level (how well examinees did on the test) on effort ratings (how hard they were trying). However, to the extent that ImPACT test performance is correlated with paper-and-pencil test performance, it is still possible that the examiners' MBC ratings were influenced by athletes' performance on paper-and-pencil tests.

Each of these individual validity criteria has its shortcomings. However, taken as a whole, our validity analysis provides tentative support for the utility of this measure. Future studies that experimentally manipulate incentives (e.g., small monetary incentives for better performance) or quasiexperimentally manipulate the circumstance of testing (e.g., testing in a group vs. individual setting or testing after practice/workout) would provide further support for the validity of MBC.

Raters had observed the test-taker's behavior over the course of a 1.5-hour evaluation including a variety of paper-and-pencil and computerized tests. Other research has demonstrated that motivation ratings with predictive validity can be made based on 15-min observations (Duckworth et al., 2011). However, we cannot address the extent to which the MBC is useful for shorter test batteries or those composed solely of computerized tests. The robustness of MBC ratings over shorter observation intervals is an important topic for future research. Furthermore, the proposed MBC factor structure should be confirmed in an independent sample. Of note, computerized tests were administered in an individualized testing session, and not in a group setting, as is often done in sport concussion management programs. It is possible that the group testing session may influence the psychometric properties of the MBC. This is an important topic for future research.

Because the MBC is based on observer ratings, rater bias towards groups defined by sport, gender, or ethnicity is a concern. Demographic variables are often proxies for underlying cultural and educational factors that could influence effort towards testing, neuropsychological test performance, and the validity of criteria measures (Heaton, Ryan, Grant, & Matthews, 1996; Kennepohl, Shore, Nabors, & Hanks, 2004; Manly & Echemendia, 2007). Hence, the determinants and consequences of group differences in MBC scores is a complex issue that is outside of the scope of the current paper. However, differences in the MBC across sport, gender, and ethnicity are of interest and an important topic for future studies. Finally, the purpose of the MBC is to contextualize changes in cognitive test performance between baseline and postconcussion evaluations. Hence, the best test of its utility would be a within-subjects comparison of baseline and postconcussion cognitive performance in relationship to MBC scores. This investigation is currently underway. Such studies are necessary for developing guidelines for interpreting baseline MBC scores and changes in MBC scores over time.

Baseline testing is a potentially powerful tool for neuropsychological assessment. However, the influence of differing incentives, and other contextual factors, across baseline and postinjury assessments raises concerns about the validity of pre–postinjury comparisons. Valid measures of athletes' level of motivation could provide important context that aids in the interpretation of neuropsychological data. These preliminary results suggest that the MBC is reliable and corresponds with examiner and self-ratings of effort towards testing, as well as baseline cognitive test performance. Pending further validation, this measure may be a worthwhile addition to neuropsychological assessments of sports-related concussion.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Allen, L., Conder, R., Green, P., & Cox, D. (1997). *Computerized assessment of response bias. Manual*. Durham, NC: Cognisyst.

Bailey, C. M., Echemendia, R. J., & Arnett, P. A. (2006). The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. *Journal of the International Neuropsychological Society*, *12*(4), 475–484.

Barth, J., Harvey, D., Freeman, J., & Brosheck, D. (2011). Sport as a laboratory assessment model. In F. Webbe (Ed.), *The handbook of sport neuropsychology* (pp. 75–89). New York, NY: Springer.

Beck, A. T., Steer, R. A., & Brown, G. K. (2000). *BDI-Fast Screen for medical patients: Manual*. San Antonio, TX: Psychological Corporation.

Benedict, R. H. (1997). *Brief Visual Memory Test–Revised*. Odessa, FL: Psychological Assessment Resources.

Benedict, R. H., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test–Revised: Normative data and analysis of inter-form and test-retest reliability. *Clinical Neuropsychology*, *12*(1), 43–55.

Conder, R., & Allen, L. (1999). Ecological validity of simulators v. TBI litigants—Part 1: Neuropsychological data. *Archives of Clinical Neuropsychology*, *14*(1), 95–96.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences*, *108*(19), 7716–7720.

Erdal, K. (2012). Neuropsychological testing for sports-related concussion: How athletes can sandbag their baseline testing without detection. *Archives of Clinical Neuropsychology*, *27*(5), 473–479.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286.

Heaton, R. K., Ryan, L., Grant, I., & Matthews, C. G. (1996). Demographic influences on neuropsychological test performance. In *Neuropsychological assessment of neuropsychiatric disorders* (2nd ed.). New York, NY: Oxford University press.

Hunt, T. N., Ferrara, M. S., Miller, L. S., & Macciocchi, S. (2007). The effect of effort on baseline neuropsychological test scores in high school football athletes. *Archives of Clinical Neuropsychology*, *22*(5), 615–621.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200.

Kennepohl, S., Shore, D., Nabors, N., & Hanks, R. (2004). African American acculturation and neuropsychological test performance following traumatic brain injury. *Journal of the International Neuropsychological Society*, *10*(4), 566–577.

Kirkwood, M., Randolph, C., & Yeates, K. (2009). Returning pediatric athletes to play after concussion: The evidence (or lack thereof) behind baseline neuropsychological testing. *Acta Paediatrica*, *98*(9), 1409.

Lovell, M., Collins, M., Podell, K., Powell, J., & Maroon, J. (2000). *ImPACT: Immediate post-concussion assessment and cognitive testing*. Pittsburgh, PA: NeuroHealth Systems, LLC.

Manly, J. J., & Echemendia, R. J. (2007). Race-specific norms: Using the model of hypertension to understand issues of race, culture, and education in neuropsychology. *Archives of Clinical Neuropsychology*, *22*(3), 319–325.

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvořák, J., Echemendia, R. J.,… Raftery, M. (2013). Consensus statement on concussion in sport: The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *British Journal of Sports Medicine*, *47*(5), 250–258.

Psychological Corporation. (2001). *Wechsler Test of Adult Reading*. San Antonio, TX: Author.

Rabinowitz, A. R., Merritt, V. C., & Arnett, P. A. (2015). The return-to-play incentive and the effect of motivation on neuropsychological test-performance: Implications for baseline concussion testing. *Developmental Neuropsychology*, *40*(1), 29–33.

Randolph, C. (2011). Baseline neuropsychological testing in managing sport-related concussion: Does it modify risk? *Current Sports Medicine Reports*, *10*(1), 21–26.

Reynolds, C. R. (2002). *Comprehensive trail making test (CTMT)*. Austin, TX: Pro-Ed.

Ruff, R., Light, R., Parker, S., & Levin, H. (1996). Benton Controlled Oral Word Association Test: Reliability and updated norms. *Archives of Clinical Neuropsychology*, *11*(4), 329–338.

Schatz, P. (2010). Long-term test–retest reliability of baseline cognitive assessments using ImPACT. *The American Journal of Sports Medicine*, *38*(1), 47–53.

Schatz, P., & Glatts, C. (2013). "Sandbagging" baseline test performance on ImPACT, without detection, is more difficult than it appears. *Archives of Clinical Neuropsychology*, *28*(3), 236–244.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321.

Smith, A. (1982). *Symbol Digit Modalities Test (SDMT)*. Los Angeles, CA: Western Psychological Services.

Solomon, G. S., & Haase, R. F. (2008). Biopsychosocial characteristics and neurocognitive test performance in National Football League players: An initial assessment. *Archives of Clinical Neuropsychology*, *23*(5), 563–577.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. (1989). *Stroop Neuropsychological Screening Test manual*. Odessa, FL: Psychological Assessment Resources.

Wechsler, D. (1997). *WAIS–III: Wechsler Adult Intelligence Scale*. San Antonio, TX: Psychological Corporation.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838.