"Sandbagging" Baseline Test Performance on ImPACT, Without Detection, Is More Difficult than It Appears

Philip Schatz*, Colette Glatts

Department of Psychology, Saint Joseph's University, Philadelphia, PA, USA

*Corresponding author at: Department of Psychology, Saint Joseph's University, 222 Post Hall, Philadelphia, PA 19131, USA. Tel.: +1-610-660-1804; fax: +1-610-660-1819.
E-mail address: pschatz@sju.edu (P. Schatz).

Abstract

Participants coached to display poor effort on neuropsychological tests have successfully evaded detection. Recent research has documented that 89% college athletes instructed to perform poorly on a follow-up baseline ImPACT (Immediate Post-concussion Assessment and Cognitive Testing) test were unable to bypass detection, but otherwise, sandbagging on baseline testing has not been directly studied. In an analog study intended to measure participants' ability to successfully sandbag, we compared baseline test performance in three groups of individuals, instructed: (a) to perform their best, (b) to malinger without guidance (e.g., naïve), and (c) how to malinger (e.g., coached), using ImPACT, the Medical Symptom Validity Test (MSVT), and the Balance Error Scoring System. The MSVT identified more participants in the naïve (80%) and coached (90%) groups than those automatically "flagged" by ImPACT (60% and 75%, respectively). Inclusion of additional indicators within ImPACT increased identification to 95% of naïve and 100% of coached malingerers. These results suggest that intentional "sandbagging" on baseline neurocognitive testing can be readily detected.

Keywords: Concussion; Coached versus naïve malingering; ImPACT

Introduction

The use of neurocognitive testing for the assessment of sports-related concussion was introduced in the 1980's (Alves, Rimel, & Nelson, 1987; Barth et al., 1989) and has since received considerable attention in the media and scientific literature. The use of baseline and post-concussion cognitive testing and symptom scores have been recommended by consensus experts for all athletes participating in contact sports that place them at risk of sustaining a concussion (McCrory et al., 2009).

While neuropsychologists expect that scores derived from neuropsychological tests are a valid reflection of a patient or athlete's abilities, there are numerous factors that can affect performance. It is widely accepted that psychological and neuropsychological assessments should include a measure of a test-taker's effort (Bush et al., 2005; Suhr & Gunstad, 2000), and decreased effort among patients suspected of sustaining a mild traumatic brain injury (mTBI) has been associated with poor test performance (Green, Iverson, & Allen, 1999; Green, Rohling, Lees-Haley, & Allen, 2001; Moss, Jones, Fokias, & Quinn, 2003; Suhr & Gunstad, 2002). In this regard, sports concussion researchers have expressed concern that suboptimal effort may negatively affect test performance, thus complicating the interpretation of post-concussion test data. A common suspicion has been that athletes may intentionally underreport symptoms following a concussion; some have pointed to the lack of reliability of self-reported history (Hall, Hall, & Chapman, 2005), others to a motivation to return to athletic competition (Echemendia & Cantu, 2003), and others associated underreporting symptoms with an athlete's fear of removal from a game or losing their position on the team (Lovell & Collins, 1998). Such beliefs have been supported by survey research. Nearly, 53% of high-school football players intentionally did not report having sustained a concussion (McCrea, Hammeke, Olsen, Leo, & Guskiewicz, 2004), due to players not thinking the injury was serious enough to warrant medical attention, not wanting to be removed from competition, and players not thinking they had sustained a concussion. This trend among

high-school athletes was recently corroborated at the professional level; in a poll of National Football League (NFL) players, 56% said they would intentionally hide concussion symptoms in order to keep playing (Sporting News, 2012). Athletes have even self-reported intentionally under-performing on baseline testing, which has been referred to as "sandbagging" (Lovell, 2007), so that comparisons with post-concussion data would be more favorable (Reilly, 2011). These anecdotal claims have been substantiated by physicians treating concussed players, who report that NFL players "purposely do bad on [baseline] testing to start" (Marvez, 2012).

Researchers had already cautioned that athletes may knowingly skew baseline test results in an attempt to temper findings on post-concussion testing, ultimately assisting with return-to-play decisions (Bailey, Echemendia, & Arnett, 2006). In their study, players classified as "low motivation" demonstrated greater changes in pre- to post-concussion scores on several cognitive measures. Traditionally, invalid or irregular patterns of performance on baseline neurocognitive test performance are identified by the presence of outliers on tests of subindices. The ImPACT (Immediate Post-concussion Assessment and Cognitive Testing) test, among the most widely used computer-based concussion test in North America, automatically "flags" athletes with baseline scores below predefined cutoffs on specific subscales (Schatz, Moser, Solomon, Ott, & Karpf, 2012). Erdal (2012) recently documented the utility of these "flagged" invalid baselines to identify college athletes instructed to intentionally perform more poorly on a subsequent follow-up baseline, when compared with an original baseline; results showed that 89% of athletes attempting to "sandbag" their performance were identified using built-in indices.

Researchers have instructed participants to feign specific disorders or conditions and have also instructed participants how to bypass detection on neuropsychological tests. Such "coached" malingering has been successfully detected in participants feigning cognitive (Jelicic, Ceunen, Peters, & Merckelbach, 2011) and memory (Powell, Gfeller, Hendricks, & Sharland, 2004) impairment following brain injury, as well as attention deficit hyperactivity disorder (Sollman, Ranseen, & Berry, 2010). However, the ability of "coached" malingerers to avoid detection on neurocognitive tests used for the assessment and management of concussion has not yet been studied. We sought to compare the utility of ImPACT subindices, as well as an external symptom validity measure, in identifying coached and naïve malingerers completing baseline neurocognitive evaluations.

Materials and Methods

Participants

Participants were 60 undergraduate students who were recruited from a human subjects pool and volunteered to participate in the study. Prior to recruitment, screening procedures excluded student athletes, students diagnosed with a concussion in the past 6 months, or students who had previously taken the ImPACT test.

Measures

ImPACT is a computerized neuropsychological test battery designed specifically for the assessment of sports-related concussions (Iverson, Lovell, & Collins, 2003). ImPACT uses individual test modules to measure varying aspects of cognitive functioning, in addition to a 22-item Post-Concussion Symptom Scale (Lovell & Collins, 1998; Lovell et al., 2006). These test modules contributes to five composite scores that include: verbal memory, visual memory, reaction time, processing speed, and impulse control (Table 1). The Impulse Control composite score, however, is used for the purpose of detecting poor effort and is not traditionally used as a clinical scale to measure effects of concussion (Iverson, Lovell, & Collins, 2003). The reliability (Elbin, Schatz, & Covassin, 2011; Iverson et al., 2003; Schatz, 2009) and validity (Iverson, Gaetz, Lovell, & Collins, 2005; Maerlender et al., 2010) of ImPACT are well documented.

The Medical Symptom Validity Test (MSVT; Green, 2004) was designed to differentiate between memory impaired patients and patients exaggerating or simulating impairment. The computerized assessment takes approximately 15 min to administer and is comprised of a basic memory test involving a list of individual common nouns flashed on the screen both individually and with a familiar paired word. After a delay, the administrator presents one word from the list, verbally, and then requests the appropriate paired-associate word (i.e., "ice" and then "skate"). The validity of the MSVT in detecting simulated malingering has been established, with a documented 98.4% specificity (when administered along with the Word Memory Test; Green, Montijo, & Brockhaus, 2011) and 80% specificity. The MSVT has also been shown to withstand "coaching," with 96%–100% specificity (Weinborn, Woods, Nulsen, & Leighton, 2012).

The Balance Error Scoring System (BESS) is a tool designed to aid in assessment of the effects of mTBI on static postural stability. The use of the BESS is a recommended tool for clinicians in helping to form return to play decisions and is reliable in

Table 1. The ImPACT test battery

Test Name	Neurocognitive Domain Measured
Word Memory	Verbal recognition memory (learning and retention)
Design Memory	Spatial recognition memory (learning and retention)
X's and O's	Visual working memory and cognitive speed
Symbol Match	Memory and visual-motor speed
Color Match	Impulse inhibition and visual-motor speed
Three letter memory	Verbal working memory and cognitive speed
Symptom scale	Rating of individual self-reported symptoms
Composite scores	Contributing Scores
Verbal Memory	Word Memory (immediate and delayed)
	Symbol Match memory score
	Three letters memory score
Visual Memory	Design Memory (immediate and delayed)
	X's and O's (average counted correct reaction time)
Reaction Time	X's and O's percent correct
	Symbol Match (average reaction time for correct responses)
	Color Match (average reaction time for correct response)
Visual Motor Speed	X's and O's (average correct distracters)
	Three Letters (average counted correctly)
Impulse Control	X's and O's (total interference errors)
	Color Match (total commission errors)

providing objective information to clinicians for the assessment of a mTBI (Riemann & Guskiewicz, 2000). The BESS uses three different stances (double leg stance, single leg stance, and tandem stance) on two different mediums (foam surface and firm surface). Errors are assigned to the test-taker when any of the following occur: (a) moving the hands off the iliac crest (hips), (b) opening the eyes, (c) step, stumble, or fall, (d) abduction or flexion of the hip beyond 30°, (e) lifting the toes or heel off the testing surface, or (f) remaining out of the proper testing position for more than 5 s. The maximum total number of errors for a single stance is 10, allowing a maximum total score of 30. Higher scores on the BESS indicate poorer static postural stability, and reliability has been established for the BESS (Hunt, Ferrara, Bornstein, & Baumgartner, 2009).

Procedure

Institutional Review Board approval was obtained, and subsequent informed consent was reviewed and obtained from all participants prior to their participation in the study. Participants were randomly assigned to one of three experimental groups ($N = 20$ per group) and then provided scripts to inform their goals for performance (*see* Appendix A). Participants assigned to the control group were given simple instructions about ImPACT and its use as a baseline measure for concussion testing and informed to perform to the best of their ability. Participants assigned to the naïve malingerer group were given the same script given to controls, as well as an addendum that included instructions to perform poorly on the test. Participants assigned to the coached malingerer group were given the same script given to controls, as well as an addendum that included instructions to perform poorly on the test, but in a way as to not be detected by the sensitivity of the test measures.

Participants were instructed to read their scripts carefully and were reminded that they would be enacting this role throughout the duration of testing. All participants then completed the online version of ImPACT, followed by the MSVT (including the immediate and delayed recall portions). All tests were administered to participants in an individual session. Because the MSVT requires a 10-min time delay between part 1 and part 2, the BESS was administered between these two portions.

Analyses

Between-groups differences were assessed using one-way analyses of variance (ANOVAs; for interval/ratio data) and chi-square tests for independence (for nominal data). Multivariate analysis of variance was conducted to establish overall differences between groups. Subsequent chi-square analyses were conducted to document the ability of specific cutoffs on ImPACT and MSVT scores in identifying participants in the naïve and coached malingering groups. ImPACT scores (Table 1) were

identified as "invalid" within the program, designated with the inclusion of "+ +" in the test type variable, on the basis of the following:

    (a) Word Memory Learning Percentage <69%,
    (b) Design Memory Learning Percentage <50%,
    (c) X's and O's Total Incorrect >30,
    (d) Impulse Control composite score >30,
    (e) Three Letters: Total letters correct <8.

Cases were also classified as "invalid" on the basis of:

    (a) Visual Motor Speed composite score <25,
    (b) Reaction Time composite score >0.80,
    (c) Word Memory Correct Distractors (Immediate + Delayed) <22,
    (d) Design Memory Correct Distractors (Immediate + Delayed) <16.

MSVT scores below 90% were designated as "invalid".

Results

A total of 60 participants completed all aspects of the study. There were a majority of female participants (76% women and 23% men), with no significant between-groups differences in gender—$\chi^2(3) = 0.17$, $p = .91$. Students ranged from 19 to 23 years of age ($M = 20.67$, $SD = 1.10$), with no significant between groups differences in age—$F(2,57) = 0.47$, $p = .63$ (see Table 2 for group frequencies, means, and SDs).

A multivariate ANOVA (MANOVA) was performed with the effort group as the independent variable and the ImPACT composite scores, MSVT, and BESS total scores as the dependent variables. Wilks' λ revealed a significant multivariate effect of the effort group on test performance—$F(8, 50) = 5.13$, $p < .001$; $\eta^2 = 0.45$. Univariate analyses revealed significant effects of the effort group on all dependent measures. Post hoc analyses revealed that participants in the "do best" group distinguished themselves from participants in both the other groups on all measures other than Reaction Time and Impulse Control, on which participants in the "coached" group were able, on average, to score similar to controls. However, participants in the "naïve" group were unable to distinguish themselves from those in the "coached" group in Verbal and Visual Memory, Symptom scores, and MSVT scores (see Table 3 for group means, significance levels, and effect sizes).

All participants in the "do best" group produced valid results, based on the "built-in" indicators on ImPACT. Using these indicators, ImPACT "flagged" 70% of participants in the naïve group and 65% of participants in the coached group—$\chi^2(2) = 24.6$, $p < .001$, $V = 0.45$. MSVT identified 80% of participants in the naïve group and 90% of participants in the coached group—$\chi^2(2) = 45.0$, $p < .001$, $V = 0.61$. When ImPACT data were combined with participants "flagged" for invalid performance on Word Memory Correct Distractors (Immediate and Delayed), 95% of participants in the naïve group and 100% of participants in the coached group were identified—$\chi^2(2) = 55.2$, $p < .001$, $V = 0.68$. When ImPACT data were analyzed for invalid performance on use of Design Memory Correct Distractors (Immediate and Delayed), 90% of participants in the naïve group and 95% of participants in the coached group were identified, $\chi^2(2) = 32.5$, $p < .001$, $V = 0.52$, along with 20% false positives in the control group. There was no overt or obvious cutoff on the BESS as a means of identifying invalid responses. Group percentages are presented in Table 4.

Cutoffs on the five subscales used as invalidity indicators within ImPACT, as well as three additional indicators within ImPACT were evaluated for efficacy in identifying individuals in each group (Table 5). Impulse Control composite scores

Table 2. Demographic data for effort groups

Variable	Group			F/χ^2	Sig.
	Do best	Naive	Coached		
Age	20.8 (1.1)	20.5 (1.2)	20.8 (1.1)	0.47	.63
Gender (%)					
Men	20	25	25	0.17	.91
Women	80	75	75		

Table 3. MANOVA results

Variable	Group (M [SD])			F	Sig.	η^2
	Do best	Naïve	Coached			
Verbal Memory	87.2 (8.7)	46.8 (30.6)	61.6 (8.7)	23.1	.001	0.45
Visual Memory	73.2 (12.6)	41.2 (22.4)	53.0 (10.2)	20.7	.001	0.42
Visual Motor Speed	36.1 (6.2)	17.5 (12.7)	25.4 (6.0)	22.3	.001	0.20
Reaction Time[a]	0.60 (.07)	1.2 (0.87)	0.86 (.16)	7.2	.002	0.20
Symptom Score[a]	8.5 (9.9)	58.1 (37.5)	40.6 (30.0)	15.7	.001	0.36
Impulse Control	4.0 (2.1)	33.6 (36.8)	15.1 (8.7)	9.4	.001	0.25
MSVT	98.8 (2.2)	71.8 (21.8)	69.8 (12.6)	24.6	.001	0.46
BESS[a]	17.3 (6.6)	21.3 (8.4)	25.2 (12.2)	15.7	.001	0.11

Notes: MANOVA: $F_{(8, 50)} = 5.13$, $p < .001$; Wilk's $\lambda = 0.30$; $\eta^2 = 0.451$.
[a]Lower scores indicate better performance.

Table 4. Classification of naïve and coached malingerers

Variable	Group (%)			χ^2	Sig.	V[a]
	Do best	Naïve	Coached			
ImPACT "flagged invalid"	0	70	65	24.6	.001	0.45
MSVT[b]	0	80	90	45.0	.001	0.61
ImPACT Word Mem CD[c]	0	95	100	55.2	.001	0.68
ImPACT Des Mem CD[d]	20	90	95	32.5	.001	0.52

[a]Cramer's *V*.
[b]MSVT = Medical Symptom Validity Test <90%.
[c]Word Memory Correct Distractors (Immediate + Delayed) <22.
[d]Design Memory Correct Distractors (Immediate + Delayed) <16.

Table 5. Percentages of scores falling above/below suggested validity cutoffs

	Cutoff	Group (%)		
		Do best	Naïve	Coached
Built-in "Flags"				
Verbal Memory[a]	<69	0	65	65
Visual Memory[b]	<50	0	45	20
Impulse Control[c]	>30	0	35	0
XO Total Incorrect	>30	0	35	0
3 Letters Tot. Cor.	<8	0	60	15
Additional Indicators				
Reaction Time	>0.80	0	65	65
Visual Motor Speed	<25	0	70	60
Word Memory CD[d]	<22	0	95	100
Design Mem CD[e]	<16	20	90	95

[a]Word Memory Learning Percentage.
[b]Design Memory Learning Percentage.
[c]Impulse Control composite score >30.
[d]Word Memory Correct Distractors (Immediate + Delayed).
[e]Design Memory Correct Distractors (Immediate + Delayed).

(Color Match Total Commission Errors and X's and O's Total Incorrect) identified only 35% of naïve malingerers and none of the coached malingerers. Verbal Memory, Visual Motor Speed, and Reaction Time all identified similar numbers of naïve and coached malingerers. Incorrect responses on Word Memory Correct Distractors emerged as the single most effective discriminator, identifying 95% of naïve malingerers and 100% of coached malingerers.

Surprisingly, while significant between-groups differences were identified on Total Symptom scores, there was no logical or empirical cutoff that allowed for classification of groups. Rather, using a cutoff score of 21 identified 95% of controls, but only 55% of naïve malingerers and 50% of coached malingerers.

Discussion

We confirmed that intentionally underperforming or "sandbagging" on baseline testing using ImPACT is extremely difficult to achieve without detection. Using built-in validity indicators in ImPACT, 70% of naïve malingerers and 65% of coached malingerers were detected. Using the forced-choice validity measure within the Word Memory portion of ImPACT (Word Memory Correct Distractors), 95% of naïve malingerers and 100% of coached malingerers were detected. Similarly, using the forced-choice validity measure within the Design Memory portion of ImPACT (e.g., Design Memory Correct Distractors) identified 90% of naïve and 95% of coached malingerers. However, 20% of controls performing their best also fell below the cutoff, decreasing the utility of Design Memory Correct Distractors in this regard. Overall, these results surpassed the identification rate of the MSVT, a similar but external forced-choice symptom validity test, which identified 80% of naïve malingerers and 90% of coached malingerers.

The current results support the findings of Erdal (2012), who concluded that is it difficult for athletes to intentionally perform poorly on ImPACT without detection. Our results showed that 30%–35% of participants were able to avoid detection using only the "built-in" validity indicators (when compared with 11% in Erdal's study). Although these two studies represent the only prospective research to date on malingering on ImPACT, there were significant methodological differences between the two studies that may help explain these discrepancies. Erdal tested collegiate athletes, approximately 1.5 years after completing an initial baseline, who were instructed to score more poorly on this subsequent test (when compared with their initial baseline) without reaching threshold on the validity indicators. We compared groups of college students (non-athletes) who were taking ImPACT for the first time and assigned them to specific malingering subgroups. Familiarity with the test requirements or the specific subtests may have resulted in Erdal's athletes having a higher rate of "successfully beating the validity indicators" than the students in our study. However, Erdal's "success rate" of 11% mirrors the 10% "success rate" of the MSVT for coached malingerers in our study.

We found that individuals attempting to sandbag their baseline tests (naïve and coached, alike) failed to correctly identify distractor items, when presented in a forced-choice format. The fact that failure to identify such distractors emerged as the most sensitive measure to malingering is not surprising. Many commercial symptom validity tests either incorporate or focus solely on this construct. This subscale appears to essentially have been "laying dormant" within the ImPACT data and has not been used as a validity measure. This subscale should be incorporated into the process of reviewing baseline tests for invalid patterns of performance.

Our inclusion of the BESS appears to be the first use of this measure in a malingering study. While participants in both the naïve and coached groups scored significantly worse than controls, there were significant within-groups variability, and little discriminability or clinical utility of the measure in terms of identifying types of malingerers. Similarly, "traditional" invalidity indicators within ImPACT such as elevated Impulse Control or Reaction Time composite scores did not assist in the identification of naïve versus coached malingerers. While we identified 65% of naïve and coached malingerers with Reaction Time (RT) scores, these cases had already been "flagged" by other indices. With respect to Impulse Control scores, Erdal identified 37% of malingerers in her study, while we identified 35% of naïve malingerers and 0% of coached malingerers. Since the ImPACT test has moved from the "desktop" version to the "online" version, the incidence of invalid baseline assessments (on the basis of Impulse Control scores) has decreased significantly (Schatz, Moser, Solomon, Ott, & Karpf, 2012). This change is likely due to the use of the keyboard responses rather than mouse for left/right responses, which yielded a high level of left-right confusion. So, the Impulse Control composite score has decreased utility in identifying invalid baseline assessments, as well as limited utility in identification of malingerers.

Total Symptom Scores did not lend in the identification of malingerers, perhaps due to the obvious need to deny the presence of concussion-related symptoms. Given documented variability in baseline symptom reporting (Covassin et al., 2006), even healthy athletes and experimental controls may present with a range of symptoms at the time of testing. However, based on the current results, it appears that approximately half of malingerers (naïve or coached) suppress symptom reporting, decreasing the utility of this score for the purpose of identification and classification of malingering. Overall, the current results suggest that, ancillary to the "built-in" or "expected" indices, there are better indicators within ImPACT for use in detecting athletes attempting to sandbag on baseline evaluations.

It is important to note that despite knowledge that athletes "claim" to have intentionally performed poorly on baseline testing (Marvez, 2012; Reilly, 2011), and researchers have speculated that this practice might take place (Bailey et al., 2006), only half of athletic trainers examine baseline test results for validity (Covassin, Elbin, Stiller-Ostrowski, & Kontos, 2009). Neuropsychologists and "consensus experts" in the area of sports concussion have agreed that baseline concussion tests may be administered by technicians or non-clinicians, but results should be interpreted by a neuropsychologist (Echemendia, Herring, & Bailes, 2009; McCrory et al., 2009; Moser et al., 2007). While patterns of performance may require interpretation by a neuropsychologist (e.g., athletes with a history of multiple concussions or learning disabilities),

visual inspection of ImPACT results for the presence of "Baseline + +" does not require significant time or educational training. The results of this study, along with those of Erdal (2012) suggest that inspection of baseline test results will identify 70%–90% of athletes attempting to sandbag their baseline. In addition, post-concussion test results can also identify cases where athletes self-report no concussion-related symptoms, but score quite poorly on testing (Schatz & Sandel, 2012).

Unfortunately, athletes may become "savvy" to these findings and alter their approach, making it necessary for test developers to incorporate multiple symptom validity measures, of different formats, within the same test. While the use of Reaction Time may have clinical utility as an independent measure for malingering, it is currently redundant with other indices. However, as athletes become more familiar with these tests, and begin to tailor their "approach," the utility of reaction time may increase. As the use of concussion testing continues to increase, invalidity criteria will need to be an ongoing subject of research, given that athletes will likely continue to find ways of attempting to outsmart the tests, as well as the professionals who are attempting to care for them.

This study is not without its limitations. As we did not evaluate collegiate athletes completing preseason neurocognitive assessments, data from student volunteers may not reflect the same patterns of performance. In addition, results may not generalize to high-school or professional athletes, who may have different motivations or intentions when completing baseline evaluations. It is important to note that there are number of factors which might contribute to an invalid baseline (e.g., Attention Deficit Disorder, Learning Disability), and not all invalid baselines are the result of intentional efforts on the part of the athlete to "sandbag." As the current results require validation in larger sample, we consider them to be preliminary. Despite these limitations, these results provide evidence that sandbagging on baseline testing, without detection, is much more difficult than it appears.

Conflict of interest

Dr. Schatz has received funding to study the effects of concussion in high school and collegiate athletes from the International Brain Research Foundation and the Sports Concussion Center of New Jersey. He has also served as a consultant to ImPACT Applications, Inc., in the context of analyzing normative data and establishing age- and gender-based norms. However, ImPACT Applications, Inc. had no role in the conceptualization of the study, the collection or analysis of data, the writing of the article, or the decision to submit it for publication. Ms. Glatts has no Conflict of Interest to declare.

Appendix A: Group Scripts

"Do Best" Group (e.g., Normal Controls)

In this study, you will be asked to take a computerized cognitive test that is used to document preseason, baseline abilities in athletes. This is the same test used by college and professional athletes, and their performance on this test is used as a comparison if they sustain a concussion. This test measures attention, concentration, processing speed, and working memory, which are commonly affected by concussion.

Please follow instructions carefully and complete the test to the best of your ability.

"Naïve Malingerer" Group (e.g., "Sandbag" baseline, uncoached)

In this study, you will be asked to take a computerized cognitive test that is used to document preseason, baseline abilities in athletes. This is the same test used by college and professional athletes, and their performance on this test is used as a comparison if they sustain a concussion. This test measures attention, concentration, processing speed, and working memory, which are commonly affected by concussion.

As you take the test, I would like you to assume the role of someone who is attempting to perform poorly on this baseline test. You are thinking that if your baseline test score is low, should you sustain a concussion in the future, your post-concussion scores will match your baseline.

"Coached Malingerer" Group "Sandbag" (e.g., "Sandbag" baseline, coached)

In this study, you will be asked to take a computerized cognitive test that is used to document preseason, baseline abilities in athletes. This is the same test used by college and professional athletes, and their performance on this test is used as a

comparison if they sustain a concussion. This test measures attention, concentration, processing speed, and working memory, which are commonly affected by concussion.

As you take the test, I would like you to assume the role of someone who is attempting to perform poorly on this baseline test. You are thinking that if your baseline test score is low, should you sustain a concussion in the future, your post-concussion scores will match your baseline.

However, you should not make it obvious that you are trying to score poorly. Major exaggerations, such as remembering absolutely nothing, or taking too long to respond, are easy to detect. The test is very sensitive to these types of behaviors. Therefore, if you score too low, or make it too obvious, you will be identified as someone who is "tanking their baseline," not someone who simply did not score very high.

References

Alves, W., Rimel, R. W., & Nelson, W. E. (1987). University of Virginia prospective study of football-induced minor head injury: Status report. *Clinical Journal of Sport Medicine*, *6*, 211–218.

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. T., Alves, W. M., Ryan, T. V., Macciocchi, S. N., Rimel, R. W., Jane, J. A., et al. (1989). Head injury in sports: Neuropsychological sequelae and recovery of function. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Mild head injury*. New York: Oxford Press.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, *20 (4)*, 419–426.

Covassin, T., Elbin, R. J., 3rd, Stiller-Ostrowski, J. L., & Kontos, A. P. (2009). Immediate post-concussion assessment and cognitive testing (ImPACT) practices of sports medicine professionals. *Journal of Athletic Training*, *44 (6)*, 639–644.

Covassin, T., Swanik, C. B., Sachs, M., Kendrick, Z., Schatz, P., Zillmer, E., et al. (2006). Sex differences in baseline neuropsychological function and concussion symptoms of collegiate athletes. *British Journal of Sports and Medicine*, *40 (11)*, 923–927; discussion 927.

Echemendia, R. J., & Cantu, R. C. (2003). Return to play following sports-related mild traumatic brain injury: The role for neuropsychology. *Applied Neuropsychology*, *10 (1)*, 48–55.

Echemendia, R. J., Herring, S., & Bailes, J. (2009). Who should conduct and interpret the neuropsychological assessment in sports-related concussion? *British Journal of Sports and Medicine*, *43 (Suppl. 1)*, i32– i 35.

Elbin, R. J., Schatz, P., & Covassin, T. (2011). One-year test-retest reliability of the online version of ImPACT in high school athletes. *American Journal of Sports and Medicine*.

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.

Green, P. (2004). *Medical Symptom Validity Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, *13 (10)*, 813–819.

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology*, *18 (2)*, 86–94.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M., 3rd. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15 (12)*, 1045–1060.

Hall, R. C., Hall, R. C., & Chapman, M. J. (2005). Definition, diagnosis, and forensic implications of postconcussional syndrome. *Psychosomatics*, *46 (3)*, 195–202.

Hunt, T. N., Ferrara, M. S., Bornstein, R. A., & Baumgartner, T. A. (2009). The reliability of the modified Balance Error Scoring System. *Clinical Journal of Sport Medicine*, *19 (6)*, 471–475.

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

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2003). Interpreting change on ImPACT following sport concussion. *The Clinical Neuropsychologist*, *17 (4)*, 460–467.

Jelicic, M., Ceunen, E., Peters, M. J., & Merckelbach, H. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the structured inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, *67 (9)*, 850–855.

Lovell, M. R. (2007). ImPACT version 6.0 clinical user's manual. Retrieved from http://www.impacttest.com/.

Lovell, M. R., & Collins, M. W. (1998). Neuropsychological assessment of the college football player. *Journal of Head and Trauma Rehabilitation*, *13 (2)*, 9–26.

Lovell, M. R., Iverson, G. L., Collins, M. W., Podell, K., Johnston, K. M., Pardini, D., et al. (2006). Measurement of symptoms following sports-related concussion: Reliability and normative data for the post-concussion scale. *Applied Neuropsychology*, *13 (3)*, 166–174.

Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., et al. (2010). Examination of the construct validity of ImPACT computerized test, traditional, and experimental neuropsychological measures. *The Clinical Neuropsychologist*, *24 (8)*, 1309–1325.

Marvez, A. (2012). Players may try to beat concussion tests. Retrieved December 4, 2012, from http://msn.foxsports.com/nfl/story/NFL-players-could-try-to-beat-concussion-tests-042111.

McCrea, M., Hammeke, T., Olsen, G., Leo, P., & Guskiewicz, K. (2004). Unreported concussion in high school football players: Implications for prevention. *Clinical Journal of Sport Medicine*, *14 (1)*, 13–17.

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M., et al. (2009). Consensus statement on concussion in sport - The 3rd International Conference on concussion in sport, held in Zurich, November 2008. *Journal of Clinical Neuroscience*, *16 (6)*, 755–763.

Moser, R. S., Iverson, G. L., Echemendia, R. J., Lovell, M. R., Schatz, P., Webbe, F. M., et al. (2007). Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Archives of Clinical Neuropsychology*, *22 (8)*, 909–916.

Moss, A., Jones, C., Fokias, D., & Quinn, D. (2003). The mediating effects of effort upon the relationship between head injury severity and cognitive functioning. *Brain Injury*, *17 (5)*, 377–387.

Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom- and test-coached simulators with the test of memory malingering. *Archives of Clinical Neuropsychology*, *19 (5)*, 693–702.

Reilly, R. (2011). Talking football with Archie, Peyton, Eli. Retrieved December 4, 2012, from http://sports.espn.go.com/espn/news/story?id=6430211.

Riemann, B. L., & Guskiewicz, K. M. (2000). Effects of mild head injury on postural stability as measured through clinical balance testing. *Journal of Athletic Training*, *35 (1)*, 19–25.

Schatz, P. (2009). Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *American Journal of Sports and Medicine*, *38 (1)*, 47–53.

Schatz, P., & Sandel, N. (2012). Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. *American Journal of Sports and Medicine*.

Schatz, P., Moser, R. S., Solomon, G. S., Ott, S. D., & Karpf, R. (2012). Incidence of invalid computerized baseline neurocognitive test results in high school and college students. *Journal of Athletic Training*, *47 (3)*, 289–296.

Sollman, M. J., Ranseen, J. D., & Berry, D. T. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, *22 (2)*, 325–335.

Sporting News. (2012). NFL concussion poll: 56 percent of players would hide symptoms to stay on field. Retrieved December 4, 2012, from http://aol.sportingnews.com/nfl/story/2012-11-11/nfl-concussions-hide-symptoms-sporting-news-midseason-players-poll.

Suhr, J. A., & Gunstad, J. (2002). "Diagnosis Threat": The effect of negative expectations on cognitive performance in head injury. *Journal of Clinical and Experimental Neuropsychology*, *24 (4)*, 448–457.

Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, *15 (5)*, 415–424.

Weinborn, M., Woods, S. P., Nulsen, C., & Leighton, A. (2012). The effects of coaching on the verbal and nonverbal medical symptom validity tests. *The Clinical Neuropsychologist*, *26 (5)*, 832–849.