

Invalid Performance and the ImPACT in National Collegiate Athletic Association Division I Football Players

Ashley J. Szabo, MA; Michael L. Alosco, MA; Andrew Fedor, MA;
John Gunstad, PhD

Department of Psychology, Kent State University, OH

Context: Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) is a computerized cognitive test battery commonly used for concussion evaluation. An important aspect of these procedures is baseline testing, but researchers have suggested that many users do not use validity indices to ensure adequate effort during testing. No one has examined the prevalence of invalid performance for college football players.

Objective: To examine the prevalence of invalid scores on ImPACT testing.

Design: Cross-sectional study.

Setting: National Collegiate Athletic Association Division I university.

Patients or Other Participants: A total of 159 athletes (age = 20.3 ± 1.41 years; range = 17.8–23.7 years) from a Division I collegiate football team participated.

Intervention(s): An informational intervention regarding the importance of concussion testing to promote safety was administered before testing for the most recent season.

Main Outcome Measure(s): We examined preseason ImPACT testing data across a 3-year period (total assessments = 269). Based on invalid and sandbagging indices denoted by the ImPACT manual, protocols were examined to indicate how many invalid indices each athlete had.

Results: A total of 27.9% ($n = 75$) of assessments were suggestive of invalid scores, with 4.1% ($n = 11$) suggesting invalid responding only, 17.5% ($n = 47$) indicating “sandbagging” only, and 6.3% ($n = 17$) showing both invalid and sandbagging responding. The informational intervention did not reduce the prevalence of invalid responding.

Conclusions: These findings highlight the need for further information about the ImPACT validity indices and whether they truly reflect poor effort. Future work is needed to identify practices to reliably target and reduce invalid responding.

Key Words: concussions, neuropsychological tests, malin-
gering

Key Points

- More than 25% ($n = 75$) of the participants' baseline cognitive testing suggested suboptimal effort.
- Providing information about concussion risks and the importance of cognitive testing did not reduce the prevalence of poor effort; however, instructing athletes to repeat testing did reduce it.
- More research is needed to determine effective and efficient strategies to reduce the prevalence of invalid performance on Immediate Post-Concussion Assessment and Cognitive Testing.

Sport-related concussions have been a growing topic of interest in the popular media and clinical neuropsychology.¹ Concussions, often termed *mild traumatic brain injuries*, are defined as traumatically induced, typically reversible impairments of neurologic function. Researchers^{2–5} have estimated the prevalence of US sport-related concussions at about 300 000 per year. This figure is likely an underestimate because many concussions are not reported.⁶ Given its popularity and perceived risk of concussions, football is the most frequently examined sport, and an estimated 5% to 9% of collegiate players sustain concussions each year.^{2,4,7} Across various college sports, concussions compose 5.8% to 6.2% of all reported injuries.^{3,7}

Several computerized cognitive test batteries have been developed to assist in the diagnosis of sport-related concussions.^{8–10} Nearly 95% of athletic trainers in a recent survey reported conducting baseline cognitive testing of

some form.¹¹ However, only 51.9% of those athletic trainers who administered computerized baseline testing examined the tests for validity concerns.¹¹ Without valid baseline testing, determining whether postinjury testing shows a meaningful change from previous levels of functioning can be challenging.¹² For example, Hunt et al¹³ showed 11% of high school athletes given a brief neuropsychological test battery exhibited poor effort on testing and athletes with invalid protocols had lower scores on several tests within the battery.

A common computerized neuropsychological test is the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT), which is used not only by the National Football League and National Hockey League but by several US Olympic teams and many universities throughout the United States and Canada.¹² The ImPACT manual¹² provides indices to detect invalid responding during testing, but no one has examined the prevalence of invalid test

profiles within a sample of student-athletes tested under typical conditions. Data from student volunteers suggest that 25.7% exhibit suboptimal effort on testing.¹⁴ Therefore, the purpose of our study was to determine the prevalence of invalid responding in student-athletes by examining preseason ImpACT testing in football players across a 3-year period. We hypothesized that poor effort would be common within this sample because researchers^{13,14} have shown that 11% of high school students and more than 25% of student volunteers exhibited suboptimal effort on neuropsychological testing. Furthermore, athletes may be motivated to intentionally suppress their baseline performances to limit detection of concussions during the season and to be allowed to continue to participate.

METHODS

Participants

Data were collected from 159 National College Athletic Association Division I football players (age = 20.3 ± 1.41 years; range = 17.8–23.7 years) at a midsized public university during a 3-year period (total preseason assessments = 269). A total of 73 athletes (27.1%) reported having experienced at least 1 concussion. Participants provided written informed consent, and the study was approved by the Kent State University Institutional Review Board.

Measures

The ImpACT is a computerized neuropsychological test battery aimed at assessing multiple areas of cognitive functioning, including attention, concentration, memory, processing speed, reaction time, and concussion-related symptom reporting. Upon completion, 5 composite scores, including verbal memory, visual memory, visual motor speed, reaction time, and impulse control, and a total symptom score are generated.¹² Reliability studies have produced mixed results for these indices, with intraclass correlation coefficient estimates ranging from 0.15 to 0.61 in a test-retest study¹⁴ in which researchers examined baseline to 45 days and 45 days to 50 days postbaseline. In another test-retest study,¹⁵ investigators found intraclass correlation coefficient estimates ranging from 0.42 to 0.74 over 2 years. When compared with traditional neuropsychological measures, processing speed and reaction time measures have moderate correlations with similar traditional neuropsychological measures.^{16,17} Schatz et al¹⁸ examined a score that consisted of the visual memory, processing speed, and impulse control composite scores and found the sensitivity of ImpACT was 81.9% and the specificity was 89.4% for identifying individuals with concussions.

The ImpACT clinical interpretation manual¹² describes 2 types of invalid protocols: invalid profiles and “sandbagging” profiles. Invalid profiles meet predetermined criteria based primarily on composite scores that indicate an athlete has not performed to his or her true level, so the results are inaccurate (Table 1). These profiles may be invalid because the participant did not read or understand directions, had attention or learning problems, was fatigued, was distracted, or had left-right confusion. Several ranges of scores are given for invalid assessments, including low percentages

Table 1. Invalid Responses for Each Test Measure (N = 269)

Index	Invalid Responses	Invalid, % (No.)
Sandbagging		
Reaction time composite	0.80–1.5 s	2.60 (7)
Verbal memory composite	<70% correct	9.29 (25)
Visual memory composite	<60% correct	17.10 (46)
Invalid		
X's and O's	>30 total incorrect	7.06 (19)
Impulse control composite	>30	8.55 (23)
Word memory learning	<69%	1.12 (3)
Design memory learning	<50%	2.97 (8)
3 letters	<8 total letters correct	0.37 (1)

for learning on memory tests, high incorrect or low correct scores for 2 subtests, and high impulse control composite scores. More specific ranges also are given for sandbagging scores, a specific subtype of invalid profile that denotes feigning weakness. Sandbagging indicates an athlete intentionally is suppressing his or her performance with the likely intent to hide any impairment when comparing baseline scores with postconcussion assessment. Sandbagging scores are based on low verbal and visual memory composite scores and slow reaction time composite scores. According to the ImpACT manual,¹² reaction time composite scores in this range generally fall below the fifth percentile (Table 1). Given that the ImpACT manual does not offer descriptive statistics or reliability data on these indices, we do not know how accurate these indices are in truly classifying protocols as invalid or feigned. These indices have never been compared with other formal measures of effort or malingering. Furthermore, no empirical literature is available to support that individuals denoted as sandbagging in fact were actively feigning cognitive deficits.

Procedures

Preseason assessments with ImpACT were collected over 3 seasons for 159 student-athletes. Many were assessed on more than 1 occasion (total assessments = 269). All ImpACT testing was conducted during preseason practice in group format, but individuals were separated to minimize distraction. Testing was directed by athletic trainers and monitored by a clinical neuropsychologist. For the most recent season, athletes also participated in an informal concussion information session that was conducted by an athletic trainer and consisted of a video regarding consequences of concussions and a brief discussion of the importance of baseline testing. Finally, 8 athletes with the highest number of invalid indices were instructed to retake the test to determine whether their performances would improve.

Statistical Analysis

We conducted χ^2 and independent-samples *t* tests to examine between-groups differences of valid and invalid ImpACT responders on key demographic and medical variables, including age, education, number of previous concussions, total symptom score, history of hyperactivity, history of repeating a grade, history of special education, and diagnosis of a learning disorder. Next, we used follow-up logistic regression analyses to investigate whether the

aforementioned medical and demographic variables predicted valid or invalid responding in this sample of athletes. Finally, χ^2 analyses were used to examine the prevalence of invalid ImPACT performance before and after a concussion information intervention. We used SPSS Statistics for Windows (version 19.0; IBM Corporation, Armonk, NY) to perform all statistical analyses. The α level was set at .05.

RESULTS

Prevalence of Suboptimal Effort on ImPACT Testing

Our analyses revealed that 27.9% ($n = 75$) of the ImPACT assessments were suggestive of invalid responding, with 4.1% ($n = 11$) indicating invalid responding only, 17.5% ($n = 47$) indicating sandbagging, and 6.3% ($n = 17$) indicating both invalid responding and sandbagging. Overall, 10.4% ($n = 28$) of players had profiles consistent with invalid responding, and 23.9% ($n = 64$) had profiles consistent with sandbagging. Of individuals with profiles meeting ImPACT criteria for invalid profiles, 2.6% ($n = 7$) of athletes had invalid performance on 1 index, 6.7% ($n = 18$) had 2 invalid indices, and 1.1% ($n = 3$) had 3 or more invalid indices. Of individuals with profiles meeting ImPACT criteria for sandbagging, 19% ($n = 51$) of athletes had 1 index suggestive of sandbagging, 4.5% ($n = 12$) had 2 sandbagging indices, and 0.4% ($n = 1$) had 3 or more sandbagging indices. Complete percentages of invalid scores by index are provided in Table 1. When combining both types of questionable performances (ie, invalid or sandbagging), 73.5% ($n = 198$) of assessments were fully intact, 16.4% ($n = 44$) had 1 invalid index, 6.7% ($n = 18$) had 2 invalid indices, and 4.9% ($n = 13$) had 3 or more suboptimal indices.

We also examined patterns of suboptimal effort over time within the sample. Of the 81 participants who completed the ImPACT at multiple preseason assessments, 58.0% ($n = 47$) exhibited valid performances at all assessments, 23.5% ($n = 19$) produced an invalid profile at 1 assessment, and 18.5% ($n = 15$) exhibited invalid responding at 2 or more assessments. More specifically, of the 56 athletes who completed ImPACT at 2 points, 21.4% ($n = 12$) had 1 invalid profile and 17.9% ($n = 10$) had 2 invalid profiles. Of the 25 athletes who took the ImPACT over 3 seasons, 28.0% ($n = 7$) had 1 invalid profile and 16.0% ($n = 4$) had 2 and 4.0% ($n = 1$) had 3 invalid profiles.

Factors Associated With Suboptimal Effort on ImPACT Testing

We examined a series of demographic and medical variables to identify factors that may increase the likelihood of suboptimal effort on ImPACT testing (Table 2). Individuals with histories of special education were more likely to be identified as producing invalid responses during testing (8.0% [$n = 6$] versus 2.1% [$n = 4$]; $\chi^2 = 0.02$, $P = .02$). No such differences emerged for factors such as years of education or history of hyperactivity. To clarify these findings, we conducted logistic regression with all demographic and medical variables (age, education, number of previous concussions, total symptom score, history of hyperactivity, history of repeating a grade, history of special education, and diagnosis of a learning disorder) to determine whether they could predict which athletes were

Table 2. Patient Characteristics Based on Invalid Responses

Characteristic	Invalid Responses	
	0	1+
Participants, No.	194	75
Age, mean $\pm$ SD, y	19.91 $\pm$ 1.41	19.59 $\pm$ 1.36
Education, mean $\pm$ SD, y	13.40 $\pm$ 1.24	13.28 $\pm$ 1.31
Concussions, mean $\pm$ SD, No.	0.44 $\pm$ 0.88	0.47 $\pm$ 0.98
Total symptoms, mean $\pm$ SD, No.	2.56 $\pm$ 6.00	3.93 $\pm$ 9.14
Hyperactivity problems, % (No.) ^a	9.8 (19)	8.0 (6)
Repeated a grade, % (No.) ^a	4.1 (8)	8.0 (6)
Special education, % (No.) ^{a,b}	2.1 (4)	8.0 (6)
Diagnosed learning disorder, % (No.) ^a	6.2 (12)	10.7 (8)

^a Percentage refers to percentage who endorsed.

^b Indicates groups were different ($P = .02$).

more likely to exhibit invalid responding during testing. The dependent variable looked at individuals with any invalid index compared with those with valid profiles (Table 3). The overall model was different (Nagelkerke $R^2 = 0.32$, $P = .02$); however, age appeared to be driving the model (χ^2 ($\chi^2 = 0.01$, $P = .02$), and no other variables emerged as predictors of valid responding ($P > .05$ for all). Individuals with valid protocols (age = 19.91 $\pm$ 1.41 years) on average were older than those with invalid protocols (age = 19.59 $\pm$ 1.36 years).

Strategies to Minimize Suboptimal Effort

Finally, we examined possible group and individual strategies to reduce suboptimal effort on ImPACT testing. Chi-square analyses showed no difference in the prevalence of questionable test performances when comparing the 2 seasons before a concussion information intervention with the most recent season after the intervention ($N = 269$, $\chi^2_1 = 0.04$, $P = .84$, 71.9% [$n = 193$] preintervention, 72.9% [$n = 196$] postintervention). However, 7 of the 8 athletes instructed to retest after generating suboptimal testing performance produced better profiles, with 5 that were completely valid.

DISCUSSION

Our results suggest that invalid and sandbagging performances are common on the ImPACT test because more than 25% of Division I football players at our institution produced a baseline score that suggested suboptimal effort according to the ImPACT manual. Although a brief informational session on concussion risks did not improve rates of valid responding, instructing athletes to complete the ImPACT a second time did appear to improve them. This improvement does not necessarily indicate that the individuals originally were feigning difficulties, because situational or longstanding factors may have affected these scores. Several aspects of these findings warrant brief discussion.

In our study, more than one-fourth of college athletes produced scores that were invalid on computerized cognitive testing, even when provided with information about the potential risks of concussion. Poor effort may complicate interpretation of postconcussion testing and place athletes at risk for subsequent injury because distinguishing poor test performance due to mild traumatic brain injury from suboptimal effort is difficult.^{19,20} The

Table 3. Logistic Regression Analysis Examining the Predictive Validity of the Descriptive Variables Among 262 ImPACT Protocols (Full Model)

Effect	Likelihood Ratio Tests		
	χ^2 Value	Degrees of Freedom	P Value
Intercept	<0.001	0	<.001
Repeated grade	0.22	1	.64
Hyperactivity problems	1.73	1	.19
Special education	2.99	1	.08
Learning disorder	0.28	1	.60
Age, y ^a	15.95	6	.01
Education, y	8.53	6	.20
Symptom total	33.59	23	.07
Test			
Overall model			
Likelihood ratio test	65.49	44	.02

^a Indicates difference (Nagelkerke $R^2 = 0.32$, $P = .02$).

ImPACT manual¹² provides guidelines for detecting invalid and sandbagging performances, and when an ImPACT report is printed, a notice is included if the individual produces a problematic profile. However, Covassin et al¹¹ found that just over half of computerized test examiners used the validity scales at baseline. Future studies are needed to determine the best practices for using validity indices from the ImPACT and other computerized test batteries because such work may help to decrease concerns regarding the ability of baseline testing to reduce concussion risk in athletes.²¹ Similarly, additional work is needed to more clearly delineate the validity indices on the ImPACT test battery. In its current form, the ImPACT manual does not offer a clear rationale for the formulation of the validity indices, and no one has directly examined their psychometric properties, including sensitivity and specificity. We do not know how sandbagging indices differ from the other invalid indices and the degree to which they provide distinct and useful data to test administrators.

As noted, 27.9% of ImPACT assessments indicated invalid responding in our study, which is a pattern generally consistent with rates in student volunteers (25.7%)¹⁴ but considerably higher than found in high school athletes (11%).¹³ However, this pattern is not surprising because Hunt et al¹³ used well-validated, individually administered tests, and rates of poor effort are typically lower on individually versus group-administered tests.²² Unfortunately, individually administering paper-and-pencil tests to athletes often is not practical for many programs. Even our midsized university conducts more than 200 preseason evaluations each year, and the extra resources for such evaluations may not be readily available.

Such findings suggest a strong need to identify strategies to improve the validity of computerized concussion testing. As mentioned, inspecting scores after testing would be appropriate to improve validity. A brief information session about concussion safety did not affect the prevalence of poor effort in our study. Researchers should examine whether other educational interventions may be more effective. However, such approaches may be difficult in student-athletes because knowledge of risk alone often is insufficient to produce behavioral change.^{23,24} A more effective approach may involve personally requesting student-athletes to undergo retesting, as 7 of 8 athletes

exhibiting invalid responding improved their test performance on repeat testing. Larger studies are needed to replicate this finding and examine whether the threat of repeat testing reduces the prevalence of suboptimal effort at the group level over time.

When combined with past work, our results suggest a possible approach to improve the validity of baseline ImPACT test performance. First, individuals administering the ImPACT test are encouraged to follow the recommendations of Lezak et al²⁵ to obtain valid testing data, including providing a quiet testing space, removing distractions (eg, cell phones), and informing athletes about the importance of providing optimal effort. Other improvements to the testing environment include having trained personnel monitor testing; reducing group numbers; and taking note of athlete characteristics that may influence test scores, such as illness, sleep, and stress. All test performances should be screened immediately upon completion to identify the possible presence of an invalid or sandbagging profile. Athletes should be informed of their questionable test performances, and they should repeat testing at least 24 to 48 hours later. This delay is encouraged to alleviate potential fatigue.²⁶ Given that the potential practice effects are unknown, investigators should determine the appropriate interval between testing sessions to optimize the clinical utility of these tests. If repeat testing indicates invalid or sandbagging profiles, careful discussion with the athlete may reveal factors, such as illness, sleeping problems, attention or learning problems, or psychological distress, that may have contributed to his or her performance.^{12,27} For example, student-athletes with histories of special education were more likely to be identified as demonstrating suboptimal effort on testing in our study. Therefore, invalid profiles actually may be valid results, which are lower than the norm because of the legitimately weak cognitive skills of the athlete. Depending on the outcome of this conversation, referring the athlete for a more thorough neuropsychological evaluation may be beneficial because he or she may have an undiagnosed attention or learning disability.¹² Empirical research is needed to examine whether this approach can reliably produce a higher proportion of valid scores.

CONCLUSIONS

More than 25% of college football players' baseline cognitive testing suggested suboptimal effort according to the ImPACT manual. Providing information about concussion risks and the importance of cognitive testing did not reduce the prevalence of poor effort, but instructing athletes to repeat testing did. Further research is needed to identify effective and efficient strategies that reduce the prevalence of invalid performances on the ImPACT.

ACKNOWLEDGMENTS

Indirect funding for this project was provided by grants DK075119 and HL089311 from the National Institutes of Health.

REFERENCES

- Barth JT, Broshek DK, Freeman JR. Sports: a new frontier for neuropsychology. In: Echemendia RJ, ed. *Sports Neuropsychology*. New York, NY: Guilford Press; 2006:3–17.

2. Brown CN, Guskiewicz KM, Bleiberg J. Athlete characteristics and outcome scores for computerized neuropsychological assessment: a preliminary analysis. *J Athl Train.* 2007;42(4):515–523.
3. Gessel LM, Fields SK, Collins CL, Dick RW, Comstock RD. Concussions among United States high school and collegiate athletes. *J Athl Train.* 2007;42(4):495–503.
4. Guskiewicz KM, McCrea M, Marshall SW, et al. Cumulative effects associated with recurrent concussion in collegiate football players. *JAMA.* 2003;290(19):2549–2555.
5. Randolph C, Kirkwood MW. What are the real risks of sport-related concussion, and are they modifiable? *J Int Neuropsychol Soc.* 2009;15(4):512–520.
6. Freeman JR, Barth JT, Broshek DK, Plehn K. Sports injuries. In: Silver JM, McAllister TW, Yudofsky SC, eds. *Textbook of Traumatic Brain Injury.* Washington, DC: American Psychiatric Publishing Inc; 2005:453–476.
7. Covassin T, Swanik CB, Sachs ML. Epidemiological considerations of concussions among intercollegiate athletes. *Appl Neuropsychol.* 2003;10(1):12–22.
8. Solomon GS, Haase RF. Biopsychosocial characteristics and neurocognitive test performance in National Football League players: an initial assessment. *Arch Clin Neuropsychol.* 2008;23(5):563–577.
9. Collins MW, Grindel SH, Lovell MR, et al. Relationship between concussion and neuropsychological performance in college football players. *JAMA.* 1999;282(10):964–970.
10. Hunt TN, Ferrara MS. Age-related differences in neuropsychological testing among high school athletes. *J Athl Train.* 2009;44(4):405–409.
11. Covassin T, Elbin RJ, Stiller-Ostrowski JL, Kontos AP. Immediate post-concussion assessment and cognitive testing (ImPACT) practices of sports medicine professionals. *J Athl Train.* 2009;44(6):639–644.
12. Lovell M. *ImPACT Version 6.0 Clinical Interpretation Manual.* http://www.impacttest.com/pdf/ImPACT_Clinical_Interpretation_Manual.pdf. Published 2007. Accessed February 4, 2011.
13. Hunt TN, Ferrara MS, Miller LS, Macciocchi S. The effect of effort on baseline neuropsychological test scores in high school football athletes. *Arch Clin Neuropsychol.* 2007;22(5):615–621.
14. Broglio SP, Ferrara MS, Macciocchi SN, Baumgartner TA, Elliott R. Test-retest reliability of computerized concussion assessment programs. *J Athl Train.* 2007;42(4):509–514.
15. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *Am J Sports Med.* 2010;38(1):47–53.
16. Iverson GL, Lovell MR, Collins MW. Validity of ImPACT for measuring processing speed following sports-related concussion. *J Clin Exp Neuropsychol.* 2005;27(6):683–689.
17. Schatz P, Putz BO. Cross-validation of measures used for computer-based assessment of concussion. *Appl Neuropsychol.* 2006;13(3):151–159.
18. Schatz P, Pardini JE, Lovell MR, Collins MW, Podell K. Sensitivity and specificity of the ImPACT test battery for concussion in athletes. *Arch Clin Neuropsychol.* 2006;21(1):91–99.
19. Kirkwood MW, Kirk JW. The base rate of suboptimal effort in a pediatric mild TBI sample: performance on the medical symptom validity test. *Clin Neuropsychol.* 2010;24(5):860–872.
20. Curtis KL, Greve KW, Bianchini KJ. The Wechsler Adult Intelligence Scale-III and malingering in traumatic brain injury: classification accuracy in known groups. *Assessment.* 2009;16(4):401–414.
21. Randolph C. Baseline neuropsychological testing in managing sport-related concussion: does it modify risk? *Curr Sports Med Repor.* 2011;10(1):21–26.
22. Groth-Marnat G. *Handbook of Psychological Assessment.* 4th ed. Hoboken, NJ: John Wiley & Sons Inc; 2003:66–68.
23. Ratliff-Crain J, Donald KM, Dalton J. Knowledge, beliefs, peer norms, and past behaviors as correlates of risky sexual behaviors among college students. *Psychol Health.* 1999;14(4):625–641.
24. Rimal RN, Juon H. Use of the risk perception attitude framework for promoting breast cancer prevention. *J Appl Soc Psychol.* 2010;40(2):287–310.
25. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment.* 4th ed. New York, NY: Oxford University Press; 2004:100–132.
26. Lovell M, Collins M, Iverson G, Johnston K, Bradley J. Grade 1 or “ding” concussions in high school athletes. *Am J Sports Med.* 2004;32(1):47–54.
27. Bailey C, Samples H, Broshek D, Freeman J, Barth J. The relationship between psychological distress and baseline sport-related concussion testing. *Clin J Sport Med.* 2010;20(4):272–277.

Address correspondence to John Gunstad, PhD, Department of Psychology, Kent State University, 221 Kent Hall, Kent State University, Kent, OH 44242-0001. Address e-mail to jgunstad@kent.edu.