Journal of Athletic Training 2012;47(3):289–296
doi: 10.4085/1062-6050-47.3.14

http://www.nata.org/journal-of-athletic-training

original research

Prevalence of Invalid Computerized Baseline Neurocognitive Test Results in High School and Collegiate Athletes

Philip Schatz, PhD*; Rosemarie Scolaro Moser, PhD, ABPP-RP, ABN†; Gary S. Solomon, PhD, ABN‡; Summer D. Ott, PsyD§; Robin Karpf, MD||

*Department of Psychology, Saint Joseph's University, Philadelphia, PA; †Sports Concussion Center of New Jersey, Lawrenceville; ‡Vanderbilt University School of Medicine, Nashville, TN; §Methodist Neurological Institute Concussion Center, Houston, TX; ||The Lawrenceville School, NJ

Context: Limited data are available regarding the prevalence and nature of invalid computerized baseline neurocognitive test data.

Objective: To identify the prevalence of invalid baselines on the desktop and online versions of ImPACT and to document the utility of correcting for left-right (L-R) confusion on the desktop version of ImPACT.

Design: Cross-sectional study of independent samples of high school (HS) and collegiate athletes who completed the desktop or online versions of ImPACT.

Participants or Other Participants: A total of 3769 HS (desktop = 1617, online = 2152) and 2130 collegiate (desktop = 742, online = 1388) athletes completed preseason baseline assessments.

Main Outcome Measure(s): Prevalence of 5 ImPACT validity indicators, with correction for L-R confusion (reversing left and right mouse-click responses) on the desktop version, by test version and group. Chi-square analyses were conducted for sex and attentional or learning disorders.

Results: At least 1 invalid indicator was present on 11.9% (desktop) versus 6.3% (online) of the HS baselines and 10.2% (desktop) versus 4.1% (online) of collegiate baselines; correcting for L-R confusion (desktop) decreased this overall prevalence to 8.4% (HS) and 7.5% (collegiate). Online Impulse Control scores alone yielded 0.4% (HS) and 0.9% (collegiate) invalid baselines, compared with 9.0% (HS) and 5.4% (collegiate) on the desktop version; correcting for L-R confusion (desktop) decreased the prevalence of invalid Impulse Control scores to 5.4% (HS) and 2.6% (collegiate). Male athletes and HS athletes with attention deficit or learning disorders who took the online version were more likely to have at least 1 invalid indicator. Utility of additional invalidity indicators is reported.

Conclusions: The online ImPACT version appeared to yield fewer invalid baseline results than did the desktop version. Identification of L-R confusion reduces the prevalence of invalid baselines (desktop only) and the potency of Impulse Control as a validity indicator. We advise test administrators to be vigilant in identifying invalid baseline results as part of routine concussion management and prevention programs.

Key Words: computerized testing, test validity, concussion testing, traumatic brain injuries

Key Points

- When baseline ImPACT data from high school and collegiate athletes were compared, fewer invalid results were found on the online version than on the desktop version.
- Because correction for left-right confusion on the desktop version of ImPACT reduced the number of invalid tests by nearly 50%, clinicians using the desktop version should watch for these errors and make the necessary correction.
- Personnel who administer or interpret baseline testing must be educated about and attentive to the possibility of invalid test performance.

Computerized baseline testing is widely used as a tool for diagnosing and managing sport-related concussions in high schools and universities across the country. The rationale for baseline neurocognitive testing is to increase the accuracy of return-to-play decisions by comparing an athlete's preconcussion and postconcussion neurocognitive functioning to help determine when the athlete has recovered. When postconcussion test performance is close to or better than baseline test performance and the athlete is asymptomatic with physical exertion, clinical recovery is assumed to have been achieved and the athlete is safe to return to play, provided no complicating factors are present. It is important to note that although clinical recovery may seem apparent, alterations in brain metabolism may extend beyond the time at which athletes self-report being symptom free and beyond the sensitivity of computer-based screening measures.[1,2] However, the use of preparticipation baseline neurocognitive testing has been endorsed by sport concussion experts[3] and has been shown to contribute additional valuable data and accuracy to the return-to-play decision-making process.[4]

Certified athletic trainers are often available to student-athletes or are in charge of student-athlete baseline testing programs, although only about 42% of all high schools actually have a certified athletic trainer on staff.[5] Nevertheless, sport concussion-testing software can be purchased and administered by institutional personnel who are not neurocognitive specialists. Because preseason baseline testing is thought to reflect an athlete's normal, healthy neurocognitive state, baseline data do not typically require interpretation by a neurocognitive specialist. Moreover, it is widely accepted that computer-based testing results (baseline or postconcussion) are important contributors to postconcussion decision making[3] but are not intended to be lone diagnostic measures (or markers).[6] Baseline testing in the school setting is typically conducted in groups, with a number of students tested simultaneously. Computer-based assessment is thought to decrease the time and staffing requirements that would be needed to administer and analyze a standardized battery of neurocognitive measures to an entire team of athletes.[7]

Although test administrators expect the scores obtained in neuropsychological tests to be valid measures of an athlete's performance, extraneous factors can affect performance. Test takers with mild traumatic brain injury have been shown to perform poorly on neuropsychological tests due to nervousness or fatigue[8] or intentional performance below their capabilities.[8–11] Athletes may be motivated to underreport postconcussion symptoms so they can return to competition more quickly.[12] Such beliefs have been empirically validated,[13,14] with athletes reporting fear of removal from a game or losing their position on the team or not wanting to let their teammates down. Beyond symptom underreporting, others have posited that athletes could actively underperform at baseline, thus affecting the measurement of cognitive ability at this time.[15] In this regard, an athlete's approach to baseline neurocognitive testing can be thought of as falling along a continuum, with optimal performance at the high end of the spectrum and performing below one's capabilities at the other. Similarly, with respect to postconcussion testing, given the range of possible symptoms, optimal performance could fall anywhere along the continuum. It is important to note distinctions between individuals purposefully malingering for secondary gain and athletes underperforming on baseline testing. An athlete could approach the baseline test session with a strategy to purposefully perform poorly (eg, "tank" or "sandbag" the baseline), so that postconcussion performance would compare more favorably with baseline performance. In addition to "active misrepresentation," poor performance on baseline assessments due to decreased motivation has been linked to personality factors, as well as lack of education about the need for testing.[16] In addition, suboptimal performance may also be due to environmental factors (eg, noise, distraction), confusion about test instructions, lack of interest, mechanical issues with the computer or input device, or other intraindividual or extraindividual factors.[17] Thus, it is important to check the validity of baseline test results for each athlete. In a recent survey[18] of athletic trainers from 1209 high schools, colleges, and universities regarding their application of the widely used computerized, neurocognitive concussion-assessment tool ImPACT, 95% of respondents reported using ImPACT for baseline testing but only 54.8% examined the validity of the baseline test.

The scientific literature on ImPACT reveals few published data on the rate of invalid test results. Surprisingly few concussion studies that used ImPACT (or other computerized neurocognitive tests platforms) documented the percentage of test performances that were discarded from statistical analysis due to invalid results. The ImPACT publishers[19] reported various criteria, or validity indicators, that can be used to determine whether baseline test results are suspect (Table 1). It is important to note these criteria are different for desktop (introduced in 2000) versus online (introduced in 2008) versions. The newer online version, automatically "flags" an invalid baseline by placing + + on the test report. Similarly, the desktop version (which many organizations and health care practices still use) denotes an invalid baseline by placing a ‡ below the test results on the clinical report, along with a statement regarding the invalidity of the baseline data. In both versions, however, specific indicators contributing to the invalid results are not identified. Furthermore, invalid baselines are often attributed to either left-right confusion (ie, reversing left and right mouse clicks on a choice reaction time task), or "sandbagging" (ie, intentionally poor

Table 1. ImPACT Battery and Composite Scores

Test Name	Neurocognitive Domain Measured
Word Memory	Word recognition memory (learning and retention)
Design Memory	Design recognition memory (learning and retention)
X's and *O*'s	Visual working memory, cognitive speed
Symbol Match	Memory, visual-motor speed
Color Match	Impulse inhibition, visual-motor speed
Three-Letters Memory	Verbal working memory, cognitive speed
Symptom Scale	Rating of individual self-reported symptoms
Composites (Desktop and Online)	Contributing scores or formula (average of scores presented)
Verbal Memory	Word Memory score: total percentage correct
	Symbol Match memory score: total correct (hidden)/9
	Three-Letters Memory: total letters correct/15
Design Memory	Design Memory: total percentage correct
	X's and *O*'s: total correct (memory)/12
Reaction Time	*X*'s and *O*'s: average counted correct reaction time (interference)
	Symbol Match: average correct reaction time (visible)/3
	Color Match: average correct reaction time
Visual Motor	*X*'s and *O*'s: total correct (interference)/4
Processing Speed	Three Letters: average counted correctly × 3
Impulse Control	*X*'s and *O*'s: total errors (interference)
	Color Match: total errors (commission)

performance on the part of the test taker),[19] but no tangible and obvious corrections are available to the clinician other than statistical reanalysis of the data or readministration of the baseline examination. Nevertheless, whether or not a test report is automatically flagged, there is concern that athletes with invalid protocols may not always be identified and, thus, may not be asked to retake the test.

One of the criteria used to identify suspect protocols appears to be more widely known and used than others: a value greater than 30 on the Impulse Control composite score, one of the ImPACT validity indicators. A review of the literature (PubMed and PsycINFO databases, 1999–2010) yielded 4 studies documenting rates of invalid baseline tests in high school, collegiate, and professional athletes. An Impulse Control score of greater than 30 was the indicator and was found in 2.5% to 8.7% of high school athletes,[17,20] 5.1% of collegiate athletes,[21] and 5.0% of professional athletes.[22] An additional study[23] identified a comparatively high 25% rate of invalid baseline results in collegiate athletes. However, in the latter study, ImPACT was 1 of 3 computerized tests administered consecutively to college students, and it was unclear whether Impulse Control scores were used as the lone indicator of invalid results; the authors[23] stated that they used the ImPACT guidelines[19] (which do not rely solely on Impulse Control scores) to determine the validity of the profile. By comparison, using traditional, paper-based measures, researchers[24] have documented invalid baseline results in 12% of high school athletes.

The purpose of our study was (1) to compare the prevalence of invalid baseline tests in athletes completing either the desktop or online version of ImPACT in group administrations, (2) to identify the benefits of correcting for left-right confusion on the desktop version, and (3) to identify the prevalence of other invalidity indicators beyond the widely used Impulse Control index score.

METHODS

Participants

Four samples participated in this study, all native English speakers, categorized according to high school versus collegiate group and desktop versus online version of ImPACT:

1. A sample of 1617 high school (HS) students (aged 13 to 18 years) completed preseason cognitive testing using ImPACT. All athletes were from a single HS in the northeastern United States, with 10.5% (n = 170) of the original sample (N = 1787) removed because they did not speak English as their primary language. All athletes completed the desktop version of ImPACT in a single computer laboratory, in groups of 16, and were supervised by the school's assistant athletic director.
2. A sample of 742 collegiate athletes (aged 18 to 22 years) completed preseason cognitive testing using ImPACT. All athletes were from a single university in the northeastern United States, with 2.2% (n = 16) of the original sample (N = 1811) removed because those individuals did not speak English as their primary language. All athletes completed the desktop version of ImPACT in groups of approximately 25, in a single computer laboratory, and were supervised by the school's sports medicine staff.
3. A sample of 2152 HS students (aged 13 to 18 years) completed preseason cognitive testing using ImPACT. All athletes attended 1 of several HSs in a single school district in the southern United States, with 6.8% (n = 156) of the original sample (N = 2308) removed because they did not speak English as their primary language. All athletes completed the online version of ImPACT in groups of 25 or more (depending on the school and size of the computer laboratory) and were supervised by a certified athletic trainer.
4. A sample of 1388 college students (aged 18 to 22 years) completed preseason cognitive testing using ImPACT. All athletes were from several colleges and universities in the eastern United States, with 3.1% (n = 44) of the original sample (N = 1432) removed because they did not speak English as their primary language. Athletes completed the online version of ImPACT in groups of approximately 25 and were supervised by a certified athletic trainer or member of the sports medicine staff.

Materials

Athletes completed either baseline testing on either the ImPACT desktop software (versions 2.0 through 6.0; Windows based, programmed in Visual FoxPro; ImPACT Applications, Inc, Pittsburgh, PA) or on the ImPACT online software (Internet based, programmed in Flash). ImPACT consists of 6 neuropsychological tests, each designed to target different aspects of cognitive functioning, including attention, memory, visual motor (processing) speed, and reaction time (Table 1). From these 6 tests, 5 separate composite scores are generated: Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control. More-thorough descriptions of the ImPACT subscales that contribute to the composite scores and the formulas for the composite scores are presented in Table 1; comprehensive descriptions are available in the literature.[25–27] Of note, the desktop version of ImPACT requires left and right mouse clicks for responses to a choice reaction-time test (*X*'s and *O*'s interference task), which often result in left-right (L-R) errors that can increase the Impulse Control score. In order to minimize L-R confusion, the online version uses keyboard responses instead of mouse clicks on those items requiring L-R responses. These L-R responses (whether by keyboard or mouse) contribute to the Reaction Time (RT), Visual Motor (Processing) Speed (VM), and Impulse Control (IC) composite scores. Otherwise, all stimuli in the online version are identical to those in the desktop version.

The IC score provides administrators with a useful measure of test validity.[19] A cutoff of 22 was introduced with version 2.0 of ImPACT[28] and was subsequently increased to 30 with version 6.0.[19] These cutoffs were determined by the test developers based on analyses of standardization data and outliers in the normative sample.[29]

Table 2. Validity Indicators for ImPACT Baseline, Desktop and Online Versions

Desktop[19]

1. Impulse Control >30 (sum of total errors on interference phase of *X*'s and *O*'s + total commission errors from color match)
2. Verbal Memory Learning <69% (average of total percentage correct on Word Memory, Symbol Match, + Three Letters)
3. Visual Memory Learning <50% (average of total percentage correct on Design Memory + *X*'s and *O*'s memory score)
4. *X*'s and *O*'s: Total Correct Interference >30
5. Three Letters: total letters correct <8

Online[28]

1. Impulse Control composite score >30
2. (Word Memory correct + Word Memory delayed correct) / 24 < 69%
3. (Design Memory correct + Design Memory delayed correct) / 24 < 50%
4. *X*'s and *O*'s: Total Correct Interference >30
5. Three Letters: total letters correct <8

Procedures

Athletes completed a baseline neurocognitive evaluation as part of their institutional requirements for participation in athletics. Permission for inclusion of data in research was obtained and approved by the institutional review boards. Athletes reported to their own institution's computer laboratory and had the test procedures explained. Invalid baseline tests were identified using the indicators listed in Table 2. Total number of invalid indicators was calculated for each athlete, using these criteria. *Left-right confusion* was defined as cases in which scores for the *X*'s and *O*'s Total Incorrect Interference were greater than 100 and scores for the *X*'s and *O*'s Total Correct Interference were less than 30. Correction for L-R confusion was conducted in accordance with instructions provided in the *ImPACT Clinical Interpretation Manual.*[19] Correction for L-R confusion on the IC composite score was achieved by replacing the *X*'s and *O*'s Total Correct Interference score with the Total Incorrect Interference score in the IC composite score formula. Correction for L-R confusion on the *X*'s and *O*'s subtest invalidity indicator was achieved by replacing the *X*'s and *O*'s Total Correct Interference score with the Total Incorrect Interference score. Finally, the prevalence of invalid VM and RT scores was identified (for the desktop version) as those cases with scores of less than 25 on the VM composite score and greater than 0.80 on the RT composite score. To determine invalid VM and RT scores for the online version, 95% confidence intervals were used (ie, 2 standard deviations) to determine cutoffs for VM (<20.4, HS; <26, collegiate) and RT scores (>0.76, HS and collegiate). In addition to identifying the overall prevalence of invalid VM and RT scores on baseline tests, we also calculated the *unique prevalence* of invalid VM and RT scores, which was defined as those individuals who had an invalid VM or RT score but no invalid score on any of the other invalidity indicators (Table 2). The prevalence of invalid baseline results was compared by sex and by self-reported diagnosis of attention deficit or learning disorder within each sample, using χ^2 analyses.

RESULTS

Prevalence of 1 or More Invalidity Indicators

On baseline tests, at least 1 invalid indicator was noted in 11.9% (n = 193) of desktop HS participants and 6.3% (n = 136) of online HS participants and in 10.2% (n = 75) of desktop collegiate participants and 4.1% (n = 57) of online collegiate participants (Table 3). For baselines completed using the desktop version, L-R confusion on the *X*'s and *O*'s subtest was identified in 3.6% (n = 58) of HS and 2.8% (n = 21) of collegiate participants. After correcting for L-R confusion on the desktop version (affecting the *X*'s and *O*'s subtest and IC composite score), 8.4% (n = 136) of desktop HS and 7.5% (n = 136) of collegiate baselines revealed at least 1 invalid indicator. Of note, no L-R confusion was identified on any baselines completed online.

After Bonferroni correction for 4 comparisons, the α level was set to $P < .0125$. Chi-square analyses revealed a greater likelihood of obtaining an invalid baseline on the desktop version than the online version in both the HS (11.9% versus 6.3%, $\chi^2_1 = 36.6$, $P = .001$) and collegiate (10.2% versus 4.1%, $\chi^2_1 = 31.1$, $P = .001$) samples. After correction for L-R confusion, χ^2 analyses revealed no greater likelihood of obtaining an invalid result on the desktop baseline than on the online version within the HS sample (8.4% versus 6.3%, $\chi^2_1 = 6.0$, $P = .015$) but revealed a greater likelihood in the collegiate sample (7.5% versus 4.1%, $\chi^2_1 = 11.4$, $P = .001$).

Prevalence of Invalid IC Scores Before and After Correction for L-R Confusion

On the desktop version, invalid IC scores were identified on 9.0% (n = 146) of HS and 5.4% (n = 40) of collegiate baseline tests. After correcting for L-R confusion on the

Table 3. Prevalence of Invalid ImPACT Results by Presence of Any Composite and Subtest Indicators, on the Desktop and Online Versions, With or Without Correction for Left-Right Confusion, in High School and Collegiate Athletes

	No. of Invalid Indicators, n (%)			
Athletes	0	1	2	3+
Desktop: without correction for left-right confusion				
High school (n = 1617)	1424 (88.1)	175 (10.8)	11 (0.7)	7 (0.4)
Collegiate (n = 742)	666 (89.8)	61 (8.2)	11 (1.6)	3 (0.3)
Desktop: with correction for left-right confusion				
High school (n = 1617)	1481 (91.5)	59 (3.7)	70 (4.3)	7 (0.4)
Collegiate (n = 742)	686 (92.5)	31 (4.2)	21 (2.8)	4 (0.5)
Online: correction for left-right confusion not needed				
High school (n = 2152)	2016 (93.7)	111 (5.2)	22 (1.0)	3 (0.1)
Collegiate (n = 1388)	1331 (95.9)	48 (3.5)	7 (0.5)	2 (0.1)

Table 4. Invalid ImPACT Results by Indicator and Desktop and Online Version in High School and Collegiate Athletes

Invalid Indicator	High School n (%)	Collegiate n (%)
	Desktop	
	(n = 1617)	(n = 742)
Impulse Control[a]	146 (9.0)	40 (5.4)
Impulse Control (correction for left-right confusion)[a]	88 (5.4)	19 (2.6)
Verbal Memory[b]	37 (2.3)	30 (4.0)
Visual Memory[c]	24 (1.5)	19 (2.6)
X's and *O*'s[d]	130 (8.1)	36 (4.9)
X's and *O*'s (correction for left-right confusion)[d]	72 (4.5)	15 (2.1)
Three Letters[e]	2 (0.1)	3 (0.4)
	Online	
	(n = 2152)	(n = 1388)
Impulse Control[a]	20 (0.9)	6 (0.4)
Word Memory[f]	8 (0.4)	8 (0.6)
Design Memory[g]	34 (1.5)	22 (1.6)
X's and *O*'s[d]	18 (0.8)	5 (0.4)
Three Letters[e]	85 (3.9)	29 (2.1)

[a] Impulse Control >30.
[b] Verbal Memory correct <69%.
[c] Visual Memory correct <50%.
[d] *X*'s and *O*'s: Total Incorrect Interference >30.
[e] Three Letters: total correct <8.
[f] Word Memory hits + Word Memory delayed correct / 24 < 0.69.
[g] Design Memory hits + Design Memory delayed correct / 24 < 0.50.

desktop version, the prevalence decreased to 5.4% (n = 88) of HS and 2.6% (n = 19) of collegiate baseline tests. These prevalences were markedly higher than the invalid scores observed on the online version: 0.9% (n = 20) of HS and 0.4% (n = 6) of collegiate baseline tests (Table 4). Before and after correcting for L-R confusion, the prevalence of invalid *X*'s and *O*'s scores on the desktop version was 8.0% (n = 130) versus 4.5% (n = 72), respectively, for the HS sample and 4.9% (n = 36) versus 2.0% (n = 15), respectively, for the collegiate sample. The prevalences of invalid *X*'s and *O*'s were also lower for the online version for both samples.

Prevalence of Invalid VM and RT Scores

On the desktop version, the prevalence of invalid VM scores was 2.2% (n = 35) of HS and 1.3% (n = 10) of

Table 5. Prevalence of Invalid Visual Motor Speed and Reaction Time Scores on ImPACT Desktop Version in High School and Collegiate Athletes

Number of Invalid Composites	High School (n = 1617)	Collegiate (n = 742)
	Invalid Scores, n (%)	
Visual Motor speed[a]	35 (2.2)	10 (1.3)
Reaction Time[b]	2 (0.1)	4 (0.5)
	Invalid Scores Not Accounted for by Other Indicators[c]	
	(n = 1617)	(n = 742)
Visual Motor speed[a]	23 (1.4)	6 (0.8)
Reaction Time[b]	1 (0.1)	1 (0.1)

[a] Visual Motor speed composite score <25.
[b] Reaction Time composite score >0.80.
[c] Prevalence of these indicators was calculated by subtracting cases with an invalid score on 1 of the 5 validity indicators (see Table 2) from the prevalence of the invalidity indicator in Table 5.

Table 6. Prevalence of Invalid Visual Motor Speed and Reaction Time Scores on ImPACT Online Version in High School and Collegiate Athletes

Number of Invalid Composites	High School (n = 2152)	Collegiate (n = 1388)
	Prevalence of Invalid Scores, n (%)	
Visual Motor speed[a]	61 (2.8)	36 (2.6)
Reaction Time[b]	66 (3.1)	28 (2.0)
	Invalid Scores Not Accounted for by Other Indicators[c]	
	(n = 2152)	(n = 1388)
Visual Motor speed[a]	43 (2.0)	24 (1.8)
Reaction Time[b]	54 (2.5)	21 (1.5)

[a] Visual Motor speed composite score <20.4 (high school) or <26 (college).
[b] Reaction Time composite score >0.76 (high school, college).
[c] Prevalence of these indicators was calculated by subtracting cases with an invalid score on 1 of the 5 validity indicators (see Table 2) from the prevalence of the invalidity indicator in Table 5.

collegiate baseline tests. After accounting for invalid scores based on any of the 5 validity indicators (ie, an athlete already had at least 1 invalid score on any of the 5 indicators in Table 2), only 1.4% (n = 23) of HS and 0.8% (n = 6) of collegiate VM scores remained (Table 5). The prevalence of invalid RT scores was 0.1% (n = 2) of HS and 0.5% (n = 4) of collegiate tests; after identifying scores based on the presence of any of the 5 validity indicators, only 0.1% (n = 1 each) of both HS and collegiate scores remained.

On the online version, the prevalence of invalid VM scores was 2.8% (n = 61) of HS and 2.6% (n = 36) of collegiate tests (Table 6). However, after removing cases with an invalid score on any of the other 5 validity indicators, only 2.0% (n = 43) of HS and 1.8% (n = 24) of collegiate tests remained. Invalid RT indicators were seen in 3.1% (n = 66) of HS and 2.0% (n = 28) of collegiate scores; after removing cases with an invalid score on any of the other 5 validity indicators, only 2.5% (n = 54) of HS and 1.5% (n = 21) of collegiate scores from the online version remained.

Invalidity Indicators by Sex and Attention Deficit or Learning Disorder

Analysis by sex yielded mixed results. Using the Bonferroni correction for multiple χ^2 comparisons, we required an α level of .0125 for statistical significance. Overall, only in the HS sample that was tested online did more male adolescents than female adolescents obtain invalid baseline tests ($\chi^2_1 = 8.47$, $P = .002$, 4.8% of males versus 1.5% of females; Table 7).

Self-report of attention deficit or learning disorder was identified in 6.7% (n = 108/1617, desktop) and 7.9% (n = 169/2153, online) of HS athletes and 8.0% (n = 59/642, desktop) and 9.0% (n = 125/1388, online) of collegiate athletes. Chi-square analyses revealed a prevalence of invalid baselines for athletes who reported a history of attention deficit or learning disorder in the HS sample only ($\chi^2_1 = 10.38$, $P = .001$). Within this sample of HS students completing ImPACT online, 6.3% (n = 136) obtained invalid baselines; those with self-reported attention deficit or learning disorder had a significantly higher likelihood of obtaining an invalid baseline (13%, n = 22/169) than those

Table 7. Percentage of Invalid Results on ImPACT, Desktop and Online Versions, by Sex, in High School and Collegiate Athletes[a]

	Invalid Results, n (%)			
	Desktop Version		Online Version	
Sex	High School (n = 1617)	Collegiate (n = 742)	High School (n = 2152)	Collegiate (n = 1388)
Male	83 (5.1)	36 (4.8)	104 (4.8)	39 (2.8)
Female	53 (3.3)	20 (2.7)	32 (1.5)	18 (1.3)
Total	136 (8.4)	56 (7.5)	136 (6.3)	57 (4.1)

[a] Desktop high school: $\chi^2_1 = 0.20$, $P = .67$. Desktop collegiate: $\chi^2_1 = 1.77$, $P = .18$. Online high school: $\chi^2_1 = 6.77$, $P = .009$. Online collegiate: $\chi^2_1 = 0.02$, $P = .88$. After Bonferroni correction for multiple χ^2 comparisons, an α level of .0125 was required for statistical significance.

without (5.7%, n = 114/1869). No differences were noted in the other 3 samples.

Of note, more invalid online baseline tests were noted in the HS athletes who reported a history of attention deficit or learning disorder ($\chi^2_1 = 10.38$, $P = .001$).

DISCUSSION

We documented the prevalence of suspect baseline neurocognitive test results, a measurement that has not been systematically available. The implications of these data are significant for the administration and application of ImPACT, a widely used tool for the assessment and management of neurocognitive effects of sport concussion. We provided the occurrence of invalid indicators for the desktop and online versions of ImPACT in both HS and college students and highlighted the importance of correcting IC scores for L-R confusion.

On the desktop version, the presence of a single invalidity indicator was 11.9% for high school students and 10.2% for college students; correction for L-R confusion on the *X*'s and *O*'s interference task decreased the prevalences to 8.4% and 7.5%, respectively. In contrast, the online version was associated with fewer invalid indicators: A single indicator was present for 6.3% of HS students and 4.1% of college students. Thus, it appears that the overall number of valid baseline scores improved on the online version. Why this is the case is not entirely clear because common individual and environmental factors that affect test performance would likely have been equally distributed across the 4 samples. One factor may be differences in the input devices between the desktop and online versions. Perhaps the keyboard input on the online version requires increased focus beyond the more customary and familiar mouse click on the desktop version. We are the first to document the prevalence of invalid responses in these 2 versions, so future researchers should further elucidate the possible factors affecting performance.

Males were more likely to have an invalid indicator but only if they were HS athletes using the online version; no other significant findings were associated with sex. Similarly, student-athletes with attention deficit or learning disorders were more likely to have an invalid indicator but only if they were HS students completing the online version. Whether these findings are spurious or a more systematic analysis of sex, attention deficit and learning disorders, and invalidity rates is warranted is unknown.

The utility of VM and RT composite scores as additional invalidity measures is not clear. Cutoff points (such as <25 on VM) for the online ImPACT test do not appear to be based on empirical data or traditional *z*-score outliers (eg, <2.0 or >2.0). Therefore, we recommend the use of empirically derived cutoff points.

We found that revisions implemented in the online version decreased the prevalence of invalid baseline tests due to extreme scores related to the most commonly used validity indicator, IC $>$ 30. Specifically, on the desktop version, the prevalence of suspect validity as a result of the IC indicator was 9.0% (HS) and 5.4% (collegiate) for the desktop version, compared with 0.9% (HS) and 0.4% (collegiate) for the online version. These differences for the online version are likely due to less L-R confusion than had been present on the mouse-driven choice RT task; test takers frequently favored their index fingers (ie, left clicking) over their middle fingers (ie, right clicking).

The current study reveals a 4% to 11% rate of invalid baseline tests among HS and collegiate athletes using the desktop and online versions of ImPACT. These percentages are small when compared with estimates of invalid neurocognitive data from patients with clinical or pathologic diagnoses in general clinical neuropsychological practice. Given that the HS and collegiate athletes in this study were considered generally healthy, however, the results are less than or equal to the 12% rate of invalid paper-and-pencil baseline tests previously reported[24] among HS football players. Nonetheless, invalid baseline scores from 5 to 10 of every 100 athletes result in a considerable need for reassessment. This, in turn, increases the time demands on the administrators as well as the athletes. In addition, failure to recognize the invalidity of test results may translate into decreased utility of these scores when compared with postconcussion performance.

With an increase in concussion awareness, litigation, and legislation,[29–33] academic institutions may be under increased pressure to provide concussion management programs that include preparticipation baseline testing. This is the model used by many professional sports teams. Yet even with easy access to computerized neurocognitive tests, will institutions ensure the proper and timely training of those who administer such tests, especially with regard to securing a valid, effortful performance from the examinee?

The value of neurocognitive testing in identifying and managing concussions has been documented empirically and cannot be underestimated. Furthermore, the advent of preparticipation baseline testing as an additional component to aid in return-to-play decisions has contributed greatly to the clinician's data-based judgment process. It should be noted, however, that the opposing viewpoint persists.[34,35] Still, the value of any neurocognitive test instrument in obtaining valid and reliable data from examinees depends on the knowledge of the persons administering and supervising the test. Unfortunately, it is not unheard of for student-athletes to be provided casual access to baseline or concussion testing in their homes without supervision. Individuals who allow athletes unsupervised access in an uncontrolled environment may be either unaware of the proper, standardized methods for neurocognitive testing or acting in a negligent manner.

Nonetheless, we believe it is essential that individuals who oversee the use of baseline testing become vigilant in identifying invalid baseline test results and making the necessary arrangements for timely retesting. All invalid baselines should be identified immediately after administration and the examinees should be retested in a timely fashion. For organizations and entities that continue to use the ImPACT desktop version, extra care and knowledge are required because the test does not automatically identify a suspect protocol. Factors that may affect a student's computerized neurocognitive baseline test performance and render invalid test results need to be studied systematically. A variety of factors affecting test performance have been posited and may include distractions in the test environment and effort or motivation. Standardization and control of the test environment are especially crucial during baseline testing because postconcussion testing is typically not performed in a group setting and is therefore less subject to the potential distractions and interruptions of the group format. The goal is to compare postconcussion test results with baseline test results, so it is important to accurately capture the athlete's best and most consistent test performances both before and after a concussion to permit appropriate comparisons. To help reduce baseline test invalidity, we recommend that test administrators exercise due diligence to determine that the athletes understand the purpose and nature of baseline testing, ensure that the athletes understand test instructions and what they have heard and read, encourage the athletes to provide a good effort, and control distractions in the test environment.

When baseline testing is invalid or when suboptimal performance persists despite the athlete's best effort, consultation with a trained neurocognitive specialist is advantageous in interpreting confusing test results. For example, athletes who have been diagnosed with an attention deficit or learning disorder may produce variable test results. In addition, conditions of long-standing L-R confusion or color blindness may significantly affect ImPACT test results, so that testing appears invalid when it is not. As noted earlier, with the online version, efforts have been made to reduce problems with L-R confusion by changing the test format. Furthermore, younger athletes (approximately 10 years old) may experience difficulties in reading comprehension and may misunderstand directions.

This study is not without its limitations. We did not compare group versus individualized administration of baseline testing or explore, in greater depth, mediating variables such as age, sex, intellectual level, presence of attention deficit or learning disorder, and level of sport. In addition, although we identified suspect invalid baseline tests using the indicators and cutoff points provided by the test developers, we had no external means of verifying that a baseline was, indeed, invalid. To this end, use of a symptom validity test or a follow-up interview with the athlete to address test performance and any contributing factors that might have affected validity should be conducted. Also, research aimed at looking more critically at the creation of and rationale for the validity indicators, with construct validity in mind, could provide more accurate data on invalidity.

This current study serves to inform those who administer or supervise baseline and concussion testing programs to be educated and vigilant about the prevalence of invalid baseline test performances. It also serves to alert test publishers to make available comprehensive test validity data that are easy to access and to advise institutions with baseline testing programs to provide proper in-service training and guidance to those who administer these tests.

REFERENCES

1. Jantzen KJ, Anderson B, Steinberg FL, Kelso JA. A prospective functional MR imaging study of mild traumatic brain injury in college football players. *AJNR Am J Neuroradiol*. 2004;25(5):738–745.
2. Vagnozzi R, Signoretti S, Cristofori L, et al. Assessment of metabolic brain damage and recovery following mild traumatic brain injury: a multicentre, proton magnetic resonance spectroscopic study in concussed patients. *Brain*. 2010;133(11):3232–3242.
3. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in sport: the 3rd International Conference on Concussion in Sport, held in Zurich, November 2008. *J Clin Neurosci*. 2009;16(6):755–763.
4. Van Kampen DA, Lovell MR, Pardini JE, Collins MW, Fu FH. The "value added" of neurocognitive testing after sports-related concussion. *Am J Sports Med*. 2006;34(10):1630–1635.
5. Mihoces G. Athletic trainers pushing for "athletic health care" in high schools. *USA Today*. June 18, 2008. http://www.usatoday.com/sports/preps/2008-06-18-hs-trainers_n.htm. Accessed May 9, 2011.
6. Echemendia RJ, Herring S, Bailes J. Who should conduct and interpret the neuropsychological assessment in sports-related concussion? *Br J Sports Med*. 2009;43(suppl 1):i32–i35.
7. Schatz P, Zillmer EA. Computer-based assessment of sports-related concussion. *Appl Neuropsychol*. 2003;10(1):42–47.
8. Suhr JA, Gunstad J. "Diagnosis threat": the effect of negative expectations on cognitive performance in head injury. *J Clin Exp Neuropsychol*. 2002;24(4):448–457.
9. Green P, Iverson GL, Allen L. Detecting malingering in head injury litigation with the Word Memory Test. *Brain Inj*. 1999;13(10):813–819.
10. Green P, Rohling ML, Lees-Haley PR, Allen LM III. Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Inj*. 2001;15(12):1045–1060.
11. Moss A, Jones C, Fokias D, Quinn D. The mediating effects of effort upon the relationship between head injury severity and cognitive functioning. *Brain Inj*. 2003;17(5):377–387.
12. Echemendia RJ, Cantu RC. Return to play following sports-related mild traumatic brain injury: the role for neuropsychology. *Appl Neuropsychol*. 2003;10(1):48–55.
13. Lovell MR, Collins MW, Maroon JC, et al. Inaccuracy of symptom reporting following concussion in athletes. *Med Sci Sports Exerc*. 2002;34(5):S298.
14. McCrea M, Hammeke T, Olsen G, Leo P, Guskiewicz K. Unreported concussion in high school football players: implications for prevention. *Clin J Sport Med*. 2004;14(1):13–17.
15. Bailey CM, Echemendia RJ, Arnett PA. The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. *J Int Neuropsychol Soc*. 2006;12(4):475–484.
16. Bailey CM, Arnett PA. Motivation and the assessment of sports-related concussion. In: Slobounov S, Sebastianelli W, eds. *Foundations of Sport-Related Brain Injuries*. Norwell, MA: Springer-Verlag; 2006:171–194.
17. Schatz P, Neidzwski K, Moser RS, Karpf R. Relationship between subjective test feedback provided by high-school athletes during computer-based assessment of baseline cognitive functioning and self-reported symptoms. *Arch Clin Neuropsychol*. 2010;25(4):285–292.
18. Covassin T, Elbin R III, Stiller-Ostrowski JL. Current sport-related concussion teaching and clinical practices of sports medicine professionals. *J Athl Train*. 2009;44(4):400–404.

19. Lovell MR. *ImPACT Version 6.0 Clinical Interpretation Manual*. http://impacttest.com/assets/pdf/2005ClinicalInterpretationManual.pdf. Accessed February 7, 2012.
20. 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.
21. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *Am J Sports Med*. 2009;38(1):47–53.
22. 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.
23. 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.
24. 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.
25. Iverson GL, Gaetz M, Lovell MR, Collins M. Relation between subjective fogginess and neuropsychological testing following concussion. *J Int Neuropsychol Soc*. 2004;10(6):1–3.
26. Lovell MR, Collins MW, Iverson GL, et al. Recovery from mild concussion in high school athletes. *J Neurosurg*. 2003;98(2):296–301.
27. Podell K. Computerized assessment of sports-related concussions. In: Lovell MR, Echemendia RJ, Barth JT, Collins MW, eds. *Traumatic Brain Injury in Sports*. Lisse, Netherlands: Swets & Zeitlinger; 2004:375–396.
28. Lovell M. *ImPACT Version 2.0 Clinical User's Manual*. www.impacttest.com/pdf/ImPACTClinicalUsersManual.pdf. Accessed February 7, 2012.
29. Concussion Treatment and Care Tools Act of 2010, HR 1347, 111th Cong (2010).
30. Larini R. Montclair High School football player's family sues over fatal brain hemorrhage. http://www.nj.com/news/index.ssf/2009/10/montclair_high_school_football.html. Accessed October 9, 2010.
31. Concussion identification, management and return-to-play: policy statement. New Jersey State Interscholastic Athletic Association Web site. http://www.njsiaa.org/NJSIAA/10ConcussionManagement.pdf. Accessed October 22, 2010.
32. Schwarz A. La Salle settles lawsuit with injured player for $7.5 million. *The New York Times*. November 30, 2010. http://www.nytimes.com/2009/12/01/sports/ncaafootball/01lasalle.html. Accessed May 9, 2011.
33. Concussion in sports: student-athlete safety and well-being. Tennessee Secondary School Athletic Association Web site. http://www.tssaa.org/concussion.pdf. Accessed January 31, 2011.
34. Randolph C. Baseline neuropsychological testing in managing sport-related concussion: does it modify risk? *Curr Sports Med Rep*. 2011;10(1):21–26.
35. Randolph C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train*. 2005;40(3):139–152.

Address correspondence to Philip Schatz, PhD, Department of Psychology, Saint Joseph's University, 222 Post Hall, Philadelphia, PA 19131. Address e-mail to pschatz@sju.edu.