Prevalence of Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator.

Christopher A. Abeare
Isabelle Messa
Brandon G. Zuccato
Bradley Merker
Laszlo Erdodi

Follow this and additional works at: https://scholarlycommons.henryford.com/behavioralhealth_articles

Recommended Citation

This Article is brought to you for free and open access by the Behavioral Health Services / Psychiatry at Henry Ford Health System Scholarly Commons. It has been accepted for inclusion in Behavioral Health Articles by an authorized administrator of Henry Ford Health System Scholarly Commons.
Prevalence of Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator

Christopher A. Abeare, PhD; Isabelle Messa, MSc; Brandon G. Zuccato, BA; Bradley Merker, PhD; Laszlo Erdodi, PhD

IMPORTANCE Estimated base rates of invalid performance on baseline testing (base rates of failure) for the management of sport-related concussion range from 6.1% to 40.0%, depending on the validity indicator used. The instability of this key measure represents a challenge in the clinical interpretation of test results that could undermine the utility of baseline testing.

OBJECTIVES To determine the prevalence of invalid performance on baseline testing and to assess whether the prevalence varies as a function of age and validity indicator.

DESIGN, SETTING, AND PARTICIPANTS This retrospective, cross-sectional study included data collected between January 1, 2012, and December 31, 2016, from a clinical referral center in the Midwestern United States. Participants included 7897 consecutively tested, equivalently proportioned male and female athletes aged 10 to 21 years, who completed baseline neurocognitive testing for the purpose of concussion management.

INTERVENTIONS Baseline assessment was conducted with the Immediate Postconcussion Assessment and Cognitive Testing (ImPACT), a computerized neurocognitive test designed for assessment of concussion.

MAIN OUTCOMES AND MEASURES Base rates of failure on published ImPACT validity indicators were compared within and across age groups. Hypotheses were developed after data collection but prior to analyses.

RESULTS Of the 7897 study participants, 4086 (51.7%) were male, mean (SD) age was 14.71 (1.78) years, 7820 (99.0%) were primarily English speaking, and the mean (SD) educational level was 8.79 (1.68) years. The base rate of failure ranged from 6.4% to 47.6% across individual indicators. Most of the sample (55.7%) failed at least 1 of 4 validity indicators. The base rate of failure varied considerably across age groups (117 of 140 [83.6%] for those aged 10 years to 14 of 48 [29.2%] for those aged 21 years), representing a risk ratio of 2.86 (95% CI, 2.60-3.16; \( \text{P} < .001 \)).

CONCLUSIONS AND RELEVANCE The results for base rate of failure were surprisingly high overall and varied widely depending on the specific validity indicator and the age of the examinee. The strong age association, with 3 of 4 participants aged 10 to 12 years failing validity indicators, suggests that the clinical interpretation and utility of baseline testing in this age group is questionable. These findings underscore the need for close scrutiny of performance validity indicators on baseline testing across age groups.
n the United States, it is estimated that between 1.1 million and 1.9 million sport- and recreation-related concussions occur annually in children aged 18 years or younger.1 This statistic, when combined with concerns about microstructural damage in the white matter tracts2 of the brain and the potential long-term effects of multiple concussions,3 is leading to a focus on the management of concussion as a public health concern. The heterogeneity in clinical presentation and natural history,4 with the absence of a reliable biomarker or other medical test to identify concussion,5 forces examiners to diagnose concussion based on the clinical assessment of medical history, balance testing, neurocognitive functioning, and postconcussion symptoms.

In an attempt to improve the diagnostic accuracy of concussion, baseline neurocognitive testing was introduced in the 1980s to assess preinjury neurocognitive functioning against which postinjury functioning could be compared.6 In contrast to the traditional methods of neuropsychological assessment in which the premorbid level of functioning is estimated and the test results are compared with normative data, baseline testing was implemented to reduce the inherent error in estimating premorbid functioning. Theoretically, this practice increases the validity of decision making because individuals serve as their own healthy controls.7 However, the utility of baseline testing compared with the use of normative data in postinjury evaluations has long been debated.8,9

The most recent Consensus Statement on Concussion in Sport10 concluded that baseline testing can be useful in the management of concussion, but there is not enough evidence to suggest that baseline testing should be mandatory. During the past decade, computerized neurocognitive testing has largely replaced traditional paper-and-pencil methods and has become widespread at all levels of sport10 because of the ease of administration and scoring, which allows for the baseline testing of groups and reduces barriers for use by a wider range of health care professionals.11 This expansion of the user base supports the need for monitoring the validity of assessment practices. Computerized baseline testing has the potential to contribute to this goal. However, computer-administered evaluations also introduce new concerns about the validity of the data they produce.

Performance validity is defined as the extent to which test-taking behavior provides an accurate reflection of the underlying cognitive ability that the instrument was designed to measure. Performance validity tests are objective measures designed to identify response sets that are unlikely to accurately reflect the true ability of the test taker. Valid performance is a basic assumption of neuropsychological testing, and it is necessary to arrive at valid and useful clinical decision making.12 Noncredible responding in neurologically intact populations can be attributed to inattentiveness, poor task comprehension, lack of incentive to perform well, or incentive to perform poorly. The influence of these factors varies as a function of age, with children and adolescents more likely to have invalid performance associated with inattentiveness, poor task comprehension, and a lack of appreciation for the importance of performing their best,13,14 whereas adults have been shown to also be influenced by external incentives.15-18

The base rate of failure (BRF), also known as the base rate of invalid performance, varies as a function of metric, population, and situational variables. Although there is limited research on the BRF during baseline testing, a systematic review by Gaudet and Weyand19 focused specifically on the Immediate Postconcussion and Cognitive Testing (ImPACT).20 The most widely used computerized neurocognitive test for the management of concussion. An estimated 75% of National Collegiate Athletic Association member schools use the ImPACT for baseline assessments,21 with BRF ranging from 2.7%22 to 27.9%23 (weighted mean, 6.1%).19 These numbers are likely underestimates, as they are based on embedded validity indicators (EVIs) that are less sensitive. These BRFs tend to be lower than those found in neurologically intact young adults who participate in academic research (BRF range, 18.3%-36.7%).24

Although children and adolescents typically pass performance validity tests designed for and normed on adults, elevated BRF in young examinees are commonly reported in the pediatric population and are typically attributed to the stage of cognitive development.13 The BRF for children in clinical settings is highly variable, ranging from 0% to 70% (weighted mean, 15.5%).12 Within studies, young children tend to have higher BRF,25,26 prompting some researchers to suggest that examinees younger than 10 years should be exempt from performance validity testing.27 Other studies found a more complex association between age and BRF that was mediated by the combined choice of performance validity tests and the cutoff used.25,26,28 Associations with age have also been found in athletes, with younger athletes (aged 10-12 years) having higher BRF on baseline testing than older athletes (aged 13-18 years).14

Four EVIs have been published for the ImPACT; 2 are included in the ImPACT Clinical Manual29 and the others were developed by independent research teams. In addition to the default ImPACT EVI that automatically flags invalid performance, the ImPACT Clinical Manual30 also provides a second EVI, “Red Flags,” as a more liberal index of suboptimal performance. Two alternative EVIs have been introduced by Schatz and Glatts40 and Higgins et al41 based on experimental malingering paradigms. These new EVIs produce higher BRFs than in the ImPACT EVIs (Table 1).
The combination of the experimental and observational evidence suggests that true BRFs could be as high as 40% in high school and collegiate athletes and even higher in younger athletes. Group-based computerized testing may be associated with a higher BRF compared with 1-on-1 evaluations because of the lack of close monitoring of test-taking behavior; however, other factors, such as adherence to standardized administration and the quality of supervision, may be better predictors of BRF than group size. The lack of consensus regarding the most appropriate EVI for determining the validity of ImPACT scores, in combination with wide discrepancies in BRFs across samples, indicators, and research designs, necessitates a direct comparison of the available EVIs in a naturalistic setting. The present study was designed to compare the 4 existing ImPACT-based EVIs across age groups in a large sample of athletes undergoing baseline testing.

### Table 1. BRF for Athletes by Age and EVI

<table>
<thead>
<tr>
<th>EVI and Component</th>
<th>Age, y</th>
<th>Cutoff</th>
<th>10 (n = 140)</th>
<th>11 (n = 142)</th>
<th>12 (n = 197)</th>
<th>13 (n = 574)</th>
<th>14 (n = 3503)</th>
<th>15 (n = 1368)</th>
<th>16 (n = 816)</th>
<th>17 (n = 388)</th>
<th>18 (n = 131)</th>
<th>19 (n = 82)</th>
<th>20 (n = 48)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default ImPACT</td>
<td>Xs and Os, total incorrect, No.</td>
<td>&gt;30</td>
<td>2.9</td>
<td>2.1</td>
<td>3.0</td>
<td>0.7</td>
<td>0.9</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
<td>0.3</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>Impulse control composite, No.</td>
<td>&gt;30</td>
<td>3.6</td>
<td>2.1</td>
<td>3.0</td>
<td>0.9</td>
<td>1.1</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
<td>0.3</td>
<td>0.8</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Word memory learning, % correct</td>
<td>&lt;69</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.4</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Design memory learning, % correct</td>
<td>&lt;50</td>
<td>0.7</td>
<td>3.5</td>
<td>1.0</td>
<td>2.4</td>
<td>2.2</td>
<td>2.9</td>
<td>3.1</td>
<td>2.6</td>
<td>2.1</td>
<td>3.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3 Letters, total letters correct, No.</td>
<td>&lt;8</td>
<td>13.6</td>
<td>6.3</td>
<td>4.1</td>
<td>4.2</td>
<td>3.5</td>
<td>3.4</td>
<td>3.1</td>
<td>1.8</td>
<td>2.1</td>
<td>3.1</td>
<td>2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>ImPACT Red Flags</td>
<td>Reaction time composite, s</td>
<td>&gt;0.80</td>
<td>29.3</td>
<td>23.2</td>
<td>7.6</td>
<td>6.6</td>
<td>4.5</td>
<td>3.1</td>
<td>2.3</td>
<td>3.1</td>
<td>2.8</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Verbal memory, No.</td>
<td>&lt;70</td>
<td>29.3</td>
<td>14.1</td>
<td>17.3</td>
<td>14.8</td>
<td>12.9</td>
<td>13.7</td>
<td>11.6</td>
<td>9.3</td>
<td>6.4</td>
<td>7.6</td>
<td>9.8</td>
<td>8.3</td>
</tr>
<tr>
<td>Visual memory, No.</td>
<td>&lt;60</td>
<td>32.9</td>
<td>34.5</td>
<td>17.8</td>
<td>19.5</td>
<td>20.3</td>
<td>24.4</td>
<td>21.7</td>
<td>17.5</td>
<td>14.4</td>
<td>16.8</td>
<td>17.1</td>
<td>12.5</td>
</tr>
<tr>
<td>Processing speed</td>
<td>&lt;25</td>
<td>52.1</td>
<td>32.4</td>
<td>14.7</td>
<td>7.5</td>
<td>6.1</td>
<td>5.1</td>
<td>2.9</td>
<td>3.7</td>
<td>3.6</td>
<td>2.3</td>
<td>0.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Schatz and Glatts, 2013</td>
<td>Reaction time, s</td>
<td>&gt;0.80</td>
<td>29.3</td>
<td>23.2</td>
<td>7.6</td>
<td>6.6</td>
<td>4.5</td>
<td>3.1</td>
<td>2.3</td>
<td>3.1</td>
<td>2.8</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Visual motor speed, No.</td>
<td>&lt;25</td>
<td>52.1</td>
<td>32.4</td>
<td>14.7</td>
<td>7.5</td>
<td>6.1</td>
<td>5.1</td>
<td>2.9</td>
<td>3.7</td>
<td>3.6</td>
<td>2.3</td>
<td>0.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Word memory CD, No.</td>
<td>&lt;22</td>
<td>25.0</td>
<td>21.8</td>
<td>17.8</td>
<td>17.2</td>
<td>20.0</td>
<td>21.5</td>
<td>20.8</td>
<td>19.5</td>
<td>15.2</td>
<td>24.4</td>
<td>14.6</td>
<td>12.5</td>
</tr>
<tr>
<td>Design memory CD, No.</td>
<td>&lt;16</td>
<td>36.4</td>
<td>32.4</td>
<td>33.5</td>
<td>35.0</td>
<td>35.3</td>
<td>41.5</td>
<td>38.6</td>
<td>38.6</td>
<td>30.9</td>
<td>29.0</td>
<td>26.8</td>
<td>16.7</td>
</tr>
<tr>
<td>Higgins et al, 2016</td>
<td>Word memory learning, %</td>
<td>≥0.90</td>
<td>10.7</td>
<td>10.6</td>
<td>7.6</td>
<td>6.1</td>
<td>6.5</td>
<td>8.0</td>
<td>8.5</td>
<td>7.9</td>
<td>4.1</td>
<td>9.9</td>
<td>4.9</td>
</tr>
<tr>
<td>Word memory delayed, % correct</td>
<td>≥81</td>
<td>26.4</td>
<td>15.5</td>
<td>24.9</td>
<td>15.7</td>
<td>19.2</td>
<td>21.6</td>
<td>21.0</td>
<td>22.6</td>
<td>13.7</td>
<td>20.6</td>
<td>19.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Memory, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design memory</td>
<td>≥64.75</td>
<td>22.9</td>
<td>20.4</td>
<td>19.3</td>
<td>21.8</td>
<td>22.3</td>
<td>25.7</td>
<td>25.6</td>
<td>21.7</td>
<td>18.8</td>
<td>18.3</td>
<td>19.5</td>
<td>16.7</td>
</tr>
<tr>
<td>XO</td>
<td>≥107.5</td>
<td>97.1</td>
<td>90.8</td>
<td>73.1</td>
<td>48.3</td>
<td>39.4</td>
<td>33.1</td>
<td>28.7</td>
<td>23.8</td>
<td>22.2</td>
<td>20.6</td>
<td>23.2</td>
<td>22.9</td>
</tr>
<tr>
<td>LRE cutoff</td>
<td>≥0.23</td>
<td>62.9</td>
<td>62.0</td>
<td>46.2</td>
<td>35.0</td>
<td>35.0</td>
<td>36.3</td>
<td>33.2</td>
<td>29.5</td>
<td>21.6</td>
<td>28.2</td>
<td>20.7</td>
<td>20.8</td>
</tr>
</tbody>
</table>

**Abbreviations:** BRF, base rate of failure; CD, correct distractors; EVI, embedded validity indicator; ImPACT, Immediate Postconcussion and Cognitive Testing; LRE, logistic regression equation.

* Xs and Os is a memory task that measures visual working memory and visual processing/visual motor speed.

* The impulse control composite is a measure of the total number of errors made on the interference phase of X’s and O’s as well as errors made on the color match task.

* Reaction time composite score is comprised of the average of the following: (1) average correct reaction time of interference stage of Xs and Os, (2) symbol match average correct reaction time visible/3, and (3) color match average correct reaction time.

* The verbal memory composite score represents the average performance on (1) word memory total percent correct, (2) symbol match total correct hidden/9 multiplied by 100, and (3) 3 letters, percent total letters correct.

* The visual memory composite score is comprised of the average of (1) design memory, total percent correct, and (2) Xs and Os (total correct memory)/12 multiplied by 100.

* Visual motor speed composite score is comprised of the average of (1) total number correct/4 during interference of Xs and Os, and (2) average counted correctly multiplied by 3 from countdown phase of 3 Letters.

* Number of correctly identified distractor items.

* Number of correctly identified incorrect distractor items.

* Hits + correct distractors/24 multiplied by 100.

* These indicators were not included in the calculation of the overall BRF; they were included in the LRE cutoff. Indicators are displayed to assist with interpretation of the LRE results.

* Delay hits + correct distractor delay items.
The default ImPACT EVI was expected to produce a BRF comparable to the 6.1% reported by Gaudet and Weyandt,23 and the ImPACT Red Flags EVI was expected to have a BRF between 20% and 30%, as reported in previous studies.23,33 Based on previous reports, we predicted that the BRF would be the highest for the EVIs introduced by Schatz and Glatts16 and Higgins et al.30 However, the paucity of research on these 2 EVIs in naturalistic settings precluded a more specific prediction. We hypothesized that the BRF on all 4 EVIs would vary across age group, producing a higher BRF in younger athletes partially because of the reliance on raw scores uncorrected for developmental changes in the underlying cognitive ability. We also predicted that BRFs would be elevated as athletes became more acculturated to playing at the more competitive levels of sport, such as the collegiate level, when the incentive to underperform would be greatest.

Methods

Participants
Participants included 7897 consecutively evaluated athletes aged 10 to 21 years who completed baseline neurocognitive testing for the management of concussion. Most of the participants were English speaking (7820 [99.0%]) and right-handed (6848 [86.7%]). The most commonly reported preexisting diagnoses were attention-deficit/hyperactivity disorder (869 participants [11.0%]), dyslexia (145 [1.8%]), and autism (27 [0.3%]). Most athletes played football (1652 [20.9%]), followed by soccer (1252 [15.9%]), volleyball (770 [9.8%]), basketball (733 [9.3%]), hockey (726 [9.2%]), and field hockey (680 [8.6%]).

Procedure
Data were collected during baseline testing for concussion between January 1, 2012, and December 31, 2016, through the concussion management program of a large hospital in the Midwestern United States. The online version of the ImPACT,24 version 2.1 (ImPACT Applications Inc), was administered to all participants. Testing lasted approximately 45 minutes and was conducted in groups of approximately 20 athletes at schools, at community centers, or in a hospital-based setting. Examiners were overseen by either an athletic trainer or a licensed clinical neuropsychologist. This study was approved by the institutional research ethics boards of Henry Ford Health System and the University of Windsor, which also waived the need for participant informed consent because of the retrospective use of deidentified data. Ethical guidelines regulating research with human participants were followed throughout the study.

Measure
The ImPACT is a computerized neurocognitive test designed for baseline and postconcussion assessment. The test includes 5 performance-based cognitive indices: reaction time, visual motor speed, impulse control, and visual and verbal memory as well as an inventory of postconcussion symptoms. One-year test-retest reliabilities are variable and range from low (intraclass correlation coefficient = 0.22) to high (0.85).35,36 Convergent validity has been established against traditional neuropsychological measures.37 The ImPACT demonstrated high sensitivity (0.82) and specificity (0.89) for concussion, with an overall classification accuracy of 86%.38

Statistical Analysis
Data were analyzed with IBM SPSS, version 24.0 (IBM Corporation). No data were missing for any variables. The BRFs were calculated for the entire sample and separately for each age group (10 through 21 years), for each of the 4 EVIs and cumulative BRF, representing the percentage of athletes who failed at least 1 of the 4 EVIs. Risk ratio (RR) was calculated between the 10-year-old group and the 21-year-old group to summarize the association of age. Mean ImPACT composite scores were calculated for each age group.

Results
Of the 7897 study participants, 4086 (51.7%) were male, the mean (SD) age was 14.71 (1.78) years, 7820 (99.0%) were primarily English speaking, and the mean (SD) educational level was 8.79 (1.68) years. Across the sample of 7897 participants, the BRF was 6.4% (505 participants) for the default ImPACT EVI, 31.8% (2509) for the ImPACT Red Flags, 34.9% (2759) for the Higgins et al10 EVI, and 55.7% (4400). Examination of the BRF by age demonstrated a strong age association (Figure). The age with the highest cumulative BRF was the 10-year-old group at 83.6% (117 of 140 participants), whereas the 21-year-old group had the lowest at 29.2% (4 of 48 participants) (RR, 2.86; 95% CI, 2.60-3.16; P < .001) (Table 1). Table 2 provides ImPACT mean raw composite scores and percentiles by age.

Discussion
Baseline cognitive testing is a common practice for concussion management programs and was introduced to enhance...
Table 2. Mean Raw ImPACT Composite Scores for Athletes by Age

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sex</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal memory</td>
<td>Male</td>
<td>75.12 (31.4)</td>
<td>79.33 (40.8)</td>
<td>80.19 (47.2)</td>
<td>80.99 (48.9)</td>
<td>80.64 (47.3)</td>
<td>80.59 (47.2)</td>
<td>81.88 (46.4)</td>
<td>82.41 (46.9)</td>
<td>85.68 (55.8)</td>
<td>83.77 (50.2)</td>
<td>85.00 (55.1)</td>
<td>84.12 (60.4)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65.21 (64.4)</td>
<td>64.73 (66.3)</td>
<td>69.85 (67.1)</td>
<td>72.90 (75.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
</tr>
<tr>
<td>Visual memory</td>
<td>Male</td>
<td>65.67 (52.6)</td>
<td>66.86 (65.4)</td>
<td>71.85 (66.6)</td>
<td>71.70 (66.8)</td>
<td>71.04 (66.9)</td>
<td>70.00 (66.0)</td>
<td>71.40 (42.9)</td>
<td>72.38 (43.3)</td>
<td>75.82 (53.0)</td>
<td>75.92 (51.7)</td>
<td>77.24 (54.4)</td>
<td>78.23 (72.0)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65.21 (64.4)</td>
<td>64.73 (66.3)</td>
<td>69.85 (67.1)</td>
<td>72.90 (75.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
<td>71.75 (73.1)</td>
</tr>
<tr>
<td>Visual motor speed</td>
<td>Male</td>
<td>23.70 (16.6)</td>
<td>26.33 (28.6)</td>
<td>30.69 (47.2)</td>
<td>32.12 (33.3)</td>
<td>33.20 (38.3)</td>
<td>34.56 (43.6)</td>
<td>35.96 (36.6)</td>
<td>37.46 (42.9)</td>
<td>39.40 (50.8)</td>
<td>38.82 (46.5)</td>
<td>39.00 (49.7)</td>
<td>39.00 (52.7)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>26.18 (22.7)</td>
<td>27.72 (30.0)</td>
<td>29.52 (43.3)</td>
<td>35.39 (33.3)</td>
<td>35.93 (38.5)</td>
<td>37.40 (37.1)</td>
<td>38.94 (44.9)</td>
<td>39.16 (45.4)</td>
<td>39.16 (47.5)</td>
<td>39.16 (47.5)</td>
<td>39.16 (47.5)</td>
<td>39.16 (47.5)</td>
</tr>
<tr>
<td>Reaction time</td>
<td>Male</td>
<td>0.78 (27.7)</td>
<td>0.74 (30.0)</td>
<td>0.67 (55.9)</td>
<td>0.64 (34.4)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>0.76 (33.5)</td>
<td>0.71 (37.4)</td>
<td>0.68 (34.4)</td>
<td>0.64 (41.8)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
<td>0.64 (41.8)</td>
<td>0.63 (45.8)</td>
</tr>
<tr>
<td>Impulse control</td>
<td>Male</td>
<td>11.78 (25.1)</td>
<td>10.23 (46.2)</td>
<td>10.31 (54.2)</td>
<td>10.34 (54.1)</td>
<td>9.36 (46.9)</td>
<td>9.36 (46.9)</td>
<td>9.36 (46.9)</td>
<td>9.36 (46.9)</td>
<td>9.36 (46.9)</td>
<td>9.36 (46.9)</td>
<td>9.36 (46.9)</td>
<td>9.36 (46.9)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8.31 (43.3)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
<td>8.66 (44.2)</td>
</tr>
<tr>
<td>No. of athletes</td>
<td>Male</td>
<td>73</td>
<td>69</td>
<td>118</td>
<td>286</td>
<td>1643</td>
<td>719</td>
<td>486</td>
<td>289</td>
<td>234</td>
<td>91</td>
<td>45</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>67</td>
<td>73</td>
<td>79</td>
<td>288</td>
<td>1960</td>
<td>134</td>
<td>134</td>
<td>134</td>
<td>134</td>
<td>134</td>
<td>134</td>
<td>134</td>
</tr>
</tbody>
</table>

The clinical utility of postinjury ImPACT data for making return-to-play decisions for athletes. However, given the reports of high and fluctuating BRFs across validity indicators and cut-offs, the validity of baseline data has become a source of concern. To empirically evaluate the legitimacy of this concern and to characterize it across development, we compared the BRF on the 4 previously published EVIs for the ImPACT. As expected for the EVIs, the BRF on the default ImPACT EVI was the lowest at 6.4%, comparable with the 6.1% rate found by Gaudet and Weyandt. The ImPACT Red Flags identified 31.8% of the sample as invalid, and the Higgins et al logistic regression equation identified 34.9%. The EVI with the highest BRF was developed by Schatz and Glatts at 47.6%. The cumulative BRF (individuals with ≥1 EVI failure) was 55.7%.

Consistent with our prediction, the youngest athletes (aged 10-11 years) had the highest BRF across all 4 EVIs. Age and invalid performance had a strong negative association; older athletes were less likely to fail EVIs. However, contrary to our prediction, the collegiate age groups did not show an increase in BRF, with the isolated exception of the 19-year-old group. The BRF continued to decrease across ages 20 and 21.

Although the literature on performance validity at baseline focuses largely on deliberate suppression of performance (ie, “sandbagging”) in high school and collegiate athletes, invalid performance is multifactorial and includes age-related fluctuations in attention and comprehension characteristic to the stage of cognitive development as well as varying ability to appreciate the importance of demonstrating one’s true ability level during cognitive testing. The very young athletes are most likely to be negatively affected by such confounding factors, which could explain the high BRF. It must be noted that these factors are likely to be independent of the conscious downward manipulation of test scores. Nevertheless, these confounding factors place examinees at higher cumulative risk of failing EVIs, particularly given that EVIs are distributed throughout the test battery as opposed to free-standing performance validity tests, which typically assess validity at discrete points in time.

The BRFs varied considerably across different EVIs. The default ImPACT EVI and ImPACT Red Flags rely on some of the same subtest scores but use different cutoffs, with default ImPACT EVI based on 2 SDs and ImPACT Red Flags based on 1.5 SDs below the mean. On the other hand, the 2 newest EVIs were calibrated empirically using experimental malingering paradigms in high school athletes and nonathlete college students. Although research suggests that performance validity tests designed for adults can be applied to children, the downward extension of adult cutoffs may violate important trajectories in cognitive development and increase the likelihood of false-positive errors in younger age groups.

Experimental malingering studies are at risk for overfitting their detection model to an atypical (ie, artificially exaggerated) manifestation of invalid responding. Namely, they were calibrated to detect the most egregious forms of invalid responding in examinees who had little incentive to avoid detection, limiting its generalizability to real-life situations with potentially substantial (perceived) reward for successful malingering. This suggests that (I) these EVIs may not apply...
Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator

Quality. Until the high BRF in young children is better understood, this group could be exempt from group-administered, computerized baseline testing. This finding raises concerns about the validity of postinjury test results in young children, suggesting that their postinjury data should be interpreted with caution. Baseline assessment of athletes aged 13 to 18 years should include procedures aimed at reducing BRF, such as strictly adhering to standardized administration practices with specific consideration to extra-test factors that may increase the BRF (limiting group size, close supervision, and minimizing distractions during testing). For the collegiate-age athletes, determinations about the validity of the baseline test scores should make use of the cumulative EVI because of the potential insensitivity of individual cutoffs in this age group.

Limitations
Providing BRF as a function of age helps clinicians contextualize the performance of athletes under their care. Knowing how common a given EVI failure is can guide diagnostic decision making because BRF foreshadows classification accuracy.40 However, the high overall BRF signals a potential confound in the measurement model. In the absence of objective, well-validated criterion measures, the degree to which the high BRF reflects false-positive errors or truly invalid response sets has yet to be determined. The standard practice of validating performance validity tests through comparison with other, well-established performance validity tests as part of a neuropsychological test battery, administered in the traditional one-on-one fashion, is warranted. This approach has the potential to isolate situational artifacts within the ImPACT (computerized group administration of a battery designed for a specific purpose). Depending on the findings, EVIs may need to be recalibrated to account for the association of examinee age and assessment context. These recommendations, combined with the complex nature of the clinical diagnosis of concussion, highlight the necessity of concussion programs to include members who have advanced knowledge of psychometric testing and performance validity assessment in addition to knowledge of management of concussion.

Conclusions
Results converge on a number of conclusions: (1) the high BRF, particularly for younger athletes, poses significant concerns regarding the validity of the baseline ImPACT data; (2) the factors that may contribute to invalid performance vary by age, suggesting that adopting age-appropriate test administration strategies (ie, item content and test instructions that match the examinee’s stage of cognitive development, individual administration, improved supervision, and strict adherence to standardized administration) may lower BRF; and (3) clinicians should routinely consider performance validity, as measured by all 4 EVIs, as well as the age-specific BRF when making return-to-play decisions based on postinjury evaluations.
Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator

REFERENCES


