ImPACT Statistics: Validity, Reliability, and Reliable Change

Statistics

Percentile Scores
Validity
Reliability
Reliable Change

NAP TIME!
STATISTICS
THE DISCIPLINE THAT PROVES
THE AVERAGE HUMAN HAS
ONE TESTICLE
The Normal Distribution

![Graph showing the normal distribution with standard deviation units (or z-scores) on the x-axis and IQ values ranging from 40 to 160 on the y-axis. Percentiles (or rounded cumulative percentages) are also shown.]
ImPACT uses percentile scores because the scores are not always normally distributed, and because percentile scores are more readily understood and interpreted than other standardized scores.
Percentile Scores and Standard Scores

Standard scores (z-scores, t-scores) are equidistant; Percentiles are NOT

+ or – 1 SD includes 68% of the sample
+ or – 2 SDs includes 96% of the sample
+ or – 3 SDs includes 99+% of the sample

+ or – 1 SD = 16th to 84th Percentiles
+ or – 2 SDs = 3rd to 98th Percentiles

For example, a percentile score of 65 means that the subject scored better than 64% of the sample (percentiles are cumulative percentages)
Is the 16th percentile abnormal? Sounds very low...but it’s low average.
ImPACT: Validity

4 studies assessing correlations between ImPACT measures and paper and pencil neuropsychological tests (Construct validity)

4 of studies assessing ImPACT’s sensitivity to concussion in athletes (Criterion validity)

Content validity is apparent in the major Composite Scores (Memory, Speed, Reaction Time)
Assessed the relationships between **Trail Making Test Part B** and the **Digit Symbol** subtest of the WAIS-III with ImPACT Visual Motor Speed (Processing Speed) Composite Score.

ImPACT correlated .53 with Digit Symbol \( (p < .003) \)
ImPACT correlated -.51 with Trails B* \( (p < .004) \)

Trails B and a test similar to Digit Symbol (SDMT) were parts of the NFL and NHL Paper and Pencil Test Batteries.

* Note: The correlation is negative because higher scores on ImPACT Visual Motor Speed reflect better performance while lower scores on Trails B reflect better performance.
Construct validity of ImPACT in athletes with concussions


**Brief Visuospatial Memory Test Total Learned** (sum 3 trials) correlated +.50 with ImPACT Verbal Memory and Visual Memory (p < .05)

**Brief Visuospatial Memory Test Delayed Recall** correlated +.85 with ImPACT Verbal Memory and Visual Memory (p < .01)

**Trail Making Test Part A** correlated -0.49 with ImPACT Visual Motor Speed) p < .05)*

**Trail Making Test part B** correlated -0.56 with ImPACT Visual Motor Speed (p < .01)*

*Note: The correlation is negative because higher scores on ImPACT Visual Motor Speed reflect better performance while lower scores on Trails A and B reflect better performance*
Validity of ImPACT for Measuring Processing Speed Following Sports-Related Concussion

GRANT L. IVESON1, MARK R. LOVELL2 AND MICHAEL W. COLLINS2


The purpose of this study was to examine the validity of ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing), a computerized neuropsychological test battery, for measuring attention and processing speed in athletes with concussions. This was accomplished by comparing the computerized testing to a traditional neuropsychological measure, the Symbol Digit Modalities Test (SDMT). Participants were 72 amateur athletes who were seen within 21 days of sustaining a sports-related concussion (Mean = 9.4, SD = 5.4 days). As predicted, the SDMT correlated more

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Memory</td>
<td>81.2</td>
<td>12.4</td>
<td>46–100</td>
<td>.46**</td>
</tr>
<tr>
<td>Visual Memory</td>
<td>72.3</td>
<td>14.9</td>
<td>29–95</td>
<td>.37**</td>
</tr>
<tr>
<td>Processing Speed</td>
<td>35.6</td>
<td>8.3</td>
<td>16.1–53.8</td>
<td>.70**</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>.58</td>
<td>.12</td>
<td>.41–1.04</td>
<td>-.60**</td>
</tr>
<tr>
<td>Total Symptoms</td>
<td>17.4</td>
<td>16.3</td>
<td>0–60</td>
<td>-.29*</td>
</tr>
<tr>
<td><strong>SDMT</strong></td>
<td>58.0</td>
<td>10.0</td>
<td>39–83</td>
<td>—</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01.
The correlations are between the ImPACT composite scores and the SDMT.
Examination of the Construct Validity of Impact™ Computerized Test, Traditional, and Experimental Neuropsychological Measures

Dartmouth Medical School, Lebanon, NH, USA.

The Clinical Neuropsychologist, 2010; 30:1-17

“In this healthy sample the correlations between the domains covered by ImPACT™ and the neuropsychological battery supports ImPACT™ as a useful screening tool for assessing many of the cognitive factors related to mTBI”.
Criterion-related Validity

Does the instrument contain the proper criteria for measuring the traits or constructs of interest?
104 high school and collegiate athletes
87% male and 80% football players
Average age was 16.11 years and average education was 9.74 years
67.3% reported no history of concussion; 32.7% reported one prior concussion, and 12.5% reported >1 prior concussion
Athletes were tested at baseline and at 2, 7, and 14 days post-concussion
72 recently concussed and 66 never concussed (control) high school athletes were tested within 72 hours of a sports-related concussion.

Athletes were from 5 states, 79.2% male, 73% football players, with average age of 16.5 years and average education of 10.4 years.

MANOVA was significant for between groups differences; ANOVAs indicated that all 6 ImPACT indexes contributed to the differences.

ImpACT sensitivity to concussion was 81.9%, and specificity was 89.4%.
The “Value Added” of Neurocognitive Testing After Sports-Related Concussion

Derk A. Van Kampen,* Mark R. Lovell,** PhD, Jamie E. Pardini,† PhD, Michael W. Collins,‡ PhD, and Freddie H. Fu,‡ MD

The American Journal of Sports Medicine,

Methods: High school and college athletes diagnosed with a concussion were tested 2 days after injury. Postinjury neurocognitive performance (Immediate Postconcussion Assessment and Cognitive Testing) and symptom (postconcussion symptom) scores were compared with preinjury (baseline) scores and with those of an age- and education-matched noninjured athlete control group. “Abnormal” test performance was determined statistically with Reliable Change Index scores.

Concussed group: 64% had increased symptoms vs. baseline
83% had at least one abnormal ImPACT score
Net yield adding ImPACT to symptoms = 19%

Control Group: 30% had one abnormal symptom or ImPACT score
0% of the control group had both abnormal symptom and ImPACT scores
Sensitivity and Specificity of the Online Version of ImPACT in High School and Collegiate Athletes

Philip Schatz,*† PhD, and Natalie Sandel,‡ BS
Investigation performed at Saint Joseph’s University, Philadelphia, Pennsylvania


81 concussed athletes tested within 3 days of concussion were matched with 81 non-concussed athletes on gender, age, sport, concussion hx, ADHD/LD

ImPACT Sensitivity = 91.4%, Specificity = 69.1%

Group of 37 concussed athletes reporting no symptoms but had invalid ImPACT tests
Excluded the subtests that contributed to invalid scores (Impulse Control > 30, Verbal Memory <69) and matched with a group of 37 controls

Sensitivity = 94.6%, Specificity = 97.3%
Neurocognitive Test Performance and Symptom Reporting in Cheerleaders with Concussions

Mark. R. Lovell, Ph.D.1, and Gary S. Solomon, Ph.D.2

Mark. R. Lovell, Ph.D.
ImPACT Applications
Pittsburgh, Pennsylvania

Gary S. Solomon, Ph.D.
Department of Neurological Surgery and the Vanderbilt Sports Concussion Center
Vanderbilt University School of Medicine
Nashville, TN

Junior and senior high school female cheerleaders (n=138) underwent pre-participation baseline testing and also repeated ImPACT within 7 days of concussive injury (range 0 to 7 days, mean= 3.9 days)

<table>
<thead>
<tr>
<th>Table I. Demographic characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Position</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
As a group (N= 138), concussed cheerleaders evaluated within seven days of injury performed poorly on ImPACT relative to their own baseline (MANOVA, F=6.5, p=.0001)

Univariate analyses indicated significant changes in Visual Memory (p=.007), Visual Motor Speed (p=.003), Reaction Time (p=.0002), and Symptom score (p=.001), but not on Verbal Memory (p=.09)

Position did not have an effect

<table>
<thead>
<tr>
<th>Position</th>
<th>Flyers</th>
<th>Baseline</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
62% of concussed cheerleaders reported an *absolute* increase in symptoms compared to their baseline ratings.

Using RCI values, 36% had an increase in symptoms.

Using RCI values, 75% had either increased symptom scores or at least one abnormal ImPACT composite score.

Of the group of cheerleaders who did not report an increase in symptoms at the time of post-injury evaluation (n=60), 33% had at least one abnormal ImPACT Composite score.
This study suggests that neurocognitive assessment provides valuable information that may not always agree with symptom self-report.

33% of cheerleaders whose symptom scores were equal to baseline had at least one abnormal cognitive score.

If complete return to cognitive and symptom baseline status is a prerequisite for return to competition, then reliance on symptom report alone may be insufficient.

Study provides empirical support for using a multimodal assessment for RTP decisions.
ImPACT composite scores have higher sensitivity than many commonly used medical tests. Numbers represent percent correct value in detecting the particular condition. A value of 100 would represent the capacity to detect a condition 100 percent of the time.

*Prostate specific antigen, **Paper and pencil neuropsychological testing.
5 studies assessing the reliability of ImPACT at one week, one month, five months, one year, and two years.
Interpreting Change on ImPACT Following Sport Concussion

Grant L. Iverson¹, Mark R. Lovell², and Michael W. Collins²

The Clinical Neuropsychologist
2003, Vol. 17, No. 4, pp. 460–467

**Test Interval: About 1 week**

56 non-concussed athletes tested an average of 5.8 days apart

Average age was 17.6 years, 64% high school, 52% male

Test-retest correlation coefficients were: Verbal Memory .70
Visual Memory .67
Visual Motor Speed .86
Reaction Time .79
Symptom Score .65
One-Month Test – Retest Reliability of the ImPACT Test Battery

Philip Schatz*, Charles S. Ferris

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

Table 1. One-month test–retest reliability

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time 1</th>
<th>Time 2</th>
<th>r</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>89.5</td>
<td>92.5</td>
<td>.66</td>
<td>0.788</td>
</tr>
<tr>
<td>SD</td>
<td>8.8</td>
<td>7.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>69.6</td>
<td>71.4</td>
<td>.43</td>
<td>0.597</td>
</tr>
<tr>
<td>SD</td>
<td>11.2</td>
<td>9.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Motor Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>39.0</td>
<td>41.8</td>
<td>.78</td>
<td>0.876</td>
</tr>
<tr>
<td>SD</td>
<td>5.8</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.61</td>
<td>0.59</td>
<td>.63</td>
<td>0.767</td>
</tr>
<tr>
<td>SD</td>
<td>0.08</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptom Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>7.4</td>
<td>9.1</td>
<td>.75</td>
<td>0.810</td>
</tr>
<tr>
<td>SD</td>
<td>9.1</td>
<td>14.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

25 undergraduates
Test Interval: About 5 months

58 collegiate football players with no reported concussions completed ImPACT at preseason, midseason, and postseason.

Repeated measures ANOVA, p < .05

No significant differences on the ImPACT scores (no effect of subconcussive hits)

Correlation coefficients not provided
One-Year Test-Retest Reliability of the Online Version of ImPACT in High School Athletes

R.J. Elbin, † PhD, Philip Schatz, ‡ PhD, and Tracey Covassin, § PhD, ATC

Investigation performed at Saint Joseph’s University, Philadelphia, Pennsylvania

369 varsity athletes; none sustained a concussion between the 2 tests

| TABLE 2 | Test-Retest Reliability | | | |
| --- | --- | --- | --- | --- | --- |
| | Time 1<sup>b</sup> | Time 2<sup>b</sup> | r | ICC | 95% CI, Lower | 95% CI, Upper |
| Verbal memory | 85.6 ± 9.1 | 86.4 ± 9.1 | .45 | .619 | .532 | .689 |
| Visual memory | 72.0 ± 12.7 | 75.5 ± 14.0 | .55 | .703 | .636 | .758 |
| Motor speed | 37.5 ± 6.7 | 39.8 ± 6.8 | .74 | .851 | .818 | .879 |
| Reaction time | .59 ± .08 | .56 ± .07 | .62 | .761 | .707 | .851 |
| Symptom scale | 4.7 ± 8.5 | 4.4 ± 8.1 | .40 | .569 | .471 | .649 |

<sup>a</sup>ICC, intraclass correlation coefficient; CI, confidence interval; UER, unbiased estimate of reliability;

<sup>b</sup>Shown as mean ± standard deviation.

<sup>c</sup>df = 368; Bonferroni-corrected alpha, P < .01.
Long-Term Test-Retest Reliability of Baseline Cognitive Assessments Using ImPACT

Phil Schatz, Ph.D.
AJSM, 2010, 38, 47-53

Test Interval: About 2 years

95 collegiate athletes (54% male, no football players) completed ImPACT at baseline and approximately 2 years later

No athlete sustained a concussion

Test-retest correlations were .68 for Reaction Time, .74 for Visual Motor Speed, .65 for Visual Memory, .46 for Verbal Memory, and .43 for the Symptom Scale

Using Reliable Change and Regression Based methodologies, 0-6% of the athletes had significant changes on the ImPACT scores, and 51% of the athletes had significant changes on the Symptom Scores
The lowest reliability value is found consistently on the...

**Symptoms Scale**

This is not surprising given the day to day variability in symptoms and the conditions under which a baseline or post-concussive symptom score might be obtained.

This is also a good reason not to rely solely on self-reported symptoms as a primary index of recovery.
118 student volunteers took 3 computerized neuropsychological programs (CogSport, HeadMinder, ImPACT), and the Memory and Concentration Test, a formal measure of effort.

**Test intervals: Baseline and 45 and 50 days later**

Invalid tests were excluded.

Reliability coefficients from baseline to Day 45 ranged from
- CogSport: .23-.65
- HeadMinder: .15-.66
- ImPACT: .15-.39

Reliability coefficients from Day 45 to Day 50 ranged from
- CogSport: .39-.66
- HeadMinder: .03-.66
- ImPACT: .39-.61
Why are the test-retest correlations so low?

The authors report that the obtained correlations were lower for all tests than in previously published studies.

They questioned the appropriateness of the Pearson correlation coefficient used in previous studies, and suggested that an intraclass correlation coefficient was a better statistical measure. One possibility for the lower correlations is the use of a different correlation coefficient.

However, there are several other possibilities for the obtained results.

This was an analogue study; subjects were student volunteers, not athletes.

Motivation was different (class credit, not athletic participation or RTP).

Although administered in counterbalanced order, students took all 4 tests consecutively.

Interference effects (different words lists, stimulus designs, etc.)
The Reliability of Medical Tests and ImPACT
Correlations at two points in time

ImPACT composite scores have higher reliability than many commonly used medical tests. Numbers represent Pearson correlations. Values above .80 are considered to be high correlations.
How do we define a reliable change (a deviation from baseline or a subsequent return to baseline) in sports concussion assessment?
Defining “return to baseline” in neurocognitive scores in sports concussion assessment

Various Initial Approaches (1990s)

1. All test scores = baseline
2. All test scores are within 10% of baseline
3. All test scores are = to or > than baseline (accounts for practice effects)

Statistically-based approaches (2000-present)

4. Reliable Change Methodology (RC)
5. Standardized Regression Based Methodology (SRB)
Reliability and Reliable Change Methodology

- Cognitive test performance can be influenced by many factors (e.g., regression to the mean, practice effects, host factors, environmental variables, and other error sources), leading to variable levels of performance in test scores. This can be a serious concern for repeated measures clinical assessment.

- The **Reliable Change Index (RCI)** is a standardized difference score assessing change from Time1 to Time 2, and allows us to circumvent (at least in part) many of these psychometric concerns.

- Any change from Time1 to Time 2 is considered **significant if the magnitude of the change is sufficiently large in proportion to the associated error variance of the test**. Error variance is calculated through accounting for test-retest reliability and variation about the mean of the test.
The Reliable Change Index (RCI)

- To calculate RCI, we need a pretest score, a posttest score, SD of the control group, and the test-retest reliability of the measure.

- We can then use formulas (see Hinton-Bayre et al., *JCEN*, 1999, 21, 70-86, or Barr et al., *ACN*, 2003, 18, 91-101, or Iverson et al., *TCN*, 2003, 17, 460-467) to calculate RCIs.

- RCIs can be based on selected confidence intervals, (e.g., 70%, 80%, 90%).

- Selection of a specific CI is based on various patient factors (e.g., age, concussion history).

- ImPACT reports scores at the 80% CI

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>p</th>
<th>SEM₁</th>
<th>SEM₂</th>
<th>Sdiff</th>
<th>Confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>p</td>
<td>SEM₁</td>
<td>SEM₂</td>
<td>Sdiff</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>88.68</td>
<td>88.84</td>
<td>.86</td>
<td>5.20</td>
<td>4.43</td>
<td>6.83</td>
<td>8.75  11.21</td>
</tr>
<tr>
<td></td>
<td>9.50</td>
<td>8.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Memory</td>
<td>78.70</td>
<td>77.48</td>
<td>.40</td>
<td>7.69</td>
<td>7.28</td>
<td>10.59</td>
<td>13.55  17.37</td>
</tr>
<tr>
<td></td>
<td>13.39</td>
<td>12.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Memory</td>
<td>.543</td>
<td>.536</td>
<td>.34</td>
<td>.04</td>
<td>.03</td>
<td>.05</td>
<td>.06     .08</td>
</tr>
<tr>
<td></td>
<td>.087</td>
<td>.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction time</td>
<td>40.54</td>
<td>42.24</td>
<td>.002</td>
<td>2.86</td>
<td>2.64</td>
<td>3.89</td>
<td>4.98  6.38</td>
</tr>
<tr>
<td></td>
<td>7.64</td>
<td>7.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing speed</td>
<td>5.23</td>
<td>5.79</td>
<td>.59</td>
<td>3.99</td>
<td>5.96</td>
<td>7.17</td>
<td>9.18   11.76</td>
</tr>
<tr>
<td>Post-Concussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>6.75</td>
<td>10.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SEM = Standard error of measurement and Sdiff = Standard error of difference
## RCI *Raw* Point Change @ 80% Confidence Interval

<table>
<thead>
<tr>
<th>Composite</th>
<th>Declined</th>
<th>Improved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Memory</td>
<td>9 points</td>
<td>9 points</td>
</tr>
<tr>
<td>Visual Memory</td>
<td>14 points</td>
<td>14 points</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>.06 seconds</td>
<td>.06 seconds</td>
</tr>
<tr>
<td>Processing Speed</td>
<td>3 points</td>
<td>7 points</td>
</tr>
<tr>
<td>Post-Concussion Scale</td>
<td>10 points</td>
<td>10 points</td>
</tr>
</tbody>
</table>
Percentages of non-concussed who show change

Table 2.

Percentages of the Healthy Sample that would be Classified as Reliably Improved or Declined

Based on the .80 and .90 Confidence Intervals.

<table>
<thead>
<tr>
<th></th>
<th>.80 Confidence interval</th>
<th>.90 Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Declined</td>
<td>Improved</td>
</tr>
<tr>
<td>Verbal Memory</td>
<td>10.7%</td>
<td>16.1%</td>
</tr>
<tr>
<td>Visual Memory</td>
<td>10.7%</td>
<td>8.9%</td>
</tr>
<tr>
<td>Reaction time</td>
<td>8.9%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Processing Speed*</td>
<td>7.1%</td>
<td>8.9%</td>
</tr>
<tr>
<td>Post-Concussion Scale</td>
<td>12.5%</td>
<td>7.1%</td>
</tr>
</tbody>
</table>
Standardized Regression Based Methodology (SRB)

- Currently recommended for research studies assessing test-retest changes.
- SRB methodology may be more accurate than RCI given its ability to correct for practice effects, regression to the mean, and the impact of baseline performance.
ENOUGH
ALREADY

Time to wake up