Authors’ reply to “Response to Mayers and Redick: ‘Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues’”

Lester B. Mayers¹ and Thomas S. Redick²

¹Division of Sports Medicine, Pace University, Pleasantville, NY, USA
²Division of Science, Indiana University-Purdue University Columbus, Columbus, IN, USA

Keywords: Rejoinder; Clinical utility; Immediate Postconcussion Assessment and Cognitive Testing; Postconcussion; Psychometric issues.

We appreciate the opportunity afforded by the Editor of the Journal to reply to the comments by Schatz, Kontos, and Elbin (2012, pp. 428–434 of this issue, hereafter abbreviated as SKE) about our analysis of the concussion assessment program Immediate Postconcussion Assessment and Cognitive Testing 2.0 (ImPACT™). We consider them to be significant contributors to the study of concussion and have carefully considered their critique.

We believe that the basic issue separating our position and that of ImPACT’s advocates is illustrated by the following quotations:

From our paper:

The proper time interval between the occurrence of a brain injury (concussion) and the time of safe return-to-play is a medical decision and should ideally be evidence based. Tests utilized to achieve this purpose should be reliable (attaining acceptable test–retest repeatability), sensitive (reliably detecting an affected patient), specific (not falsely identifying normal subjects as being affected), and have acceptable validity (sensitive to impairment when true impairment is present).


GENERAL CONSIDERATIONS

1. SKE refer to our “seemingly random quotations at the start” (p. 428) of our manuscript. These quotations were cited in order to illustrate the general acceptance of ImPACT among researchers and practitioners involved with concussion. We believe that we have then proceeded to “deconstruct” this degree of acceptance. Whether we accomplished a valid analysis of the pertinent literature to prove our viewpoint or “selectively review limited observations” (p. 428) is the object of debate here.
2. SKE dismiss a previous critical neuropsychological test review (Randolph, McCrea, & Barr, 2005) stating that “As computer-based concussion assessment and management tests became commercially available in 2000, there was an extremely narrow time frame to conceptualize and complete prospective research as required by the inclusion criteria put forth by Randolph et al.” (p. 428). However, despite this “narrow time frame,” the developers and purveyors of ImPACT were able to publish no fewer than 33 articles promoting the efficacy of their product in neuropsychological and sports medical journals during that interval (Impact Applications, Inc., 2012c). We do not believe that the premise for our article was “built on a weak scientific foundation” (p. 429) as the publications by Randolph and others (Randolph, 2011; Randolph & Kirkwood, 2009) were peer reviewed, appeared in respected journals, and are widely accepted. Moreover, the discussion of Randolph and colleagues’ work has no bearing on the information that we provide in our review.

3. Regarding the general criticism by SKE that our publication is not a “review paper” but more of an “opinion” they are correct. We did not present our analysis as a comprehensive review and certainly could not perform a meta-analysis because of the small number of studies available for each component of review in our original paper. For example, at the time our manuscript was written, there were only three publications (Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007; Iverson, Lovell, & Collins, 2003; Schatz, 2010) that reported ImPACT test–retest reliability in healthy subjects. Across these studies, Iverson, Lovell, and Collins (2003) reported Pearson’s r only, Broglio et al. (2007) reported intraclass coefficients (ICCs) only, and Schatz (2010) reported both Pearson’s r and ICCs.

4. We should have provided more detailed explanation of our “framework for identifying published articles.” We reviewed published studies of ImPACT’s psychometric properties from its website (Impact Applications, Inc., 2012c), searches of library databases for pertinent peer-reviewed articles, and references from articles publishing ImPACT data. Although we did not provide this information in our original article, researchers are able to read the studies that we referenced in order to “validate the conclusions” in our review. ImPACT is a computerized neuropsychological assessment tool that is promoted by its purveyors as “reliable and valid” for determining the time of recovery after a brain injury. Our analysis has led us to conclude that it is not useful for this purpose.

SPECIFIC CONSIDERATIONS

Test–retest reliability

1. We note that despite the repeated criticism by SKE of the “selective” and “biased” nature of our review, they indicated only two articles relevant to our discussion of ImPACT. Elbin, Schatz, and Covassin (2011) published a study in November 2011 after our article was accepted, providing additional test–retest reliability. Lau, Collins, and Lovell (2011), published in June 2011 while our article was under review, examined the utility of ImPACT to distinguish between short and protracted recovery after concussion, a topic we did not address in our analysis.

2. SKE’s criticism of Broglio et al. (2007) as “methodologically flawed” (p. 429) is unsubstantiated. For example, although the number of subjects these researchers excluded because of invalid baseline ImPACT data (29 out of 118, 25%, not 34% as stated by SKE) is higher than that typically observed, the remaining 73 college students performed extremely well. The baseline ImPACT scores of the subjects studied by Broglio et al. (2007) was numerically better or the same as those of the college students reported by Schatz (2010). In addition, their suggestion of potential proactive and reactive interference effects from performing multiple similar cognitive tests within a 1-hour session is unfounded. If this claim was true, than virtually every large-scale factor-analytic study on the relationship among working memory, short-term memory, and fluid intelligence that has been conducted would also be “methodologically flawed.” In the absence of direct evidence for why the study by Broglio et al. (2007) should be ignored, other than the fact that the ICCs they measured are lower than those found by Schatz (2010) and Elbin et al. (2011), we feel that it is necessary to include Broglio et al. (2007) in our discussion of ImPACT’s test–retest reliability (Table 1).

3. We agree that multiple measures of test–retest reliability are desirable for ImPACT, and some metrics such as reliable change indices (RCIs) and regression-based methods (RBM) are more relevant when attempting to make return-to-play (RTP) decisions about individual athletes. However, SKE fail to acknowledge the point
### Table 1
Test–retest reliability of ImPACT using Pearson *r*, intraclass correlation coefficients, and reliable change index confidence intervals

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Interval</th>
<th><em>VeM</em></th>
<th><em>ViM</em></th>
<th>PS</th>
<th>RT</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iverson, Lovell, and Collins (2003)</td>
<td>56</td>
<td>1–13 days</td>
<td>.70</td>
<td>.67</td>
<td>.86</td>
<td>.79</td>
<td>Pearson</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.21</td>
<td>17.37</td>
<td>6.38</td>
<td>0.08</td>
<td>90% RCI CIs</td>
</tr>
<tr>
<td>Broglio et al. (2007)</td>
<td>73</td>
<td>∼45 days</td>
<td>.24</td>
<td>.35</td>
<td>.40</td>
<td>.41</td>
<td>Pearson</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.23</td>
<td>.32</td>
<td>.38</td>
<td>.39</td>
<td>ICCs</td>
</tr>
<tr>
<td>Schatz (2010)</td>
<td>95</td>
<td>∼2 years</td>
<td>.30</td>
<td>.49</td>
<td>.60</td>
<td>.52</td>
<td>Pearson</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.46</td>
<td>.65</td>
<td>.74</td>
<td>.68</td>
<td>ICCs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16.08</td>
<td>20.19</td>
<td>8.71</td>
<td>0.10</td>
<td>95% RCI CIs</td>
</tr>
<tr>
<td>Elbin et al. (2011)</td>
<td>369</td>
<td>0.5–2.35 years</td>
<td>.45</td>
<td>.55</td>
<td>.74</td>
<td>.62</td>
<td>Pearson</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.62</td>
<td>.70</td>
<td>.85</td>
<td>.76</td>
<td>ICCs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15.61</td>
<td>19.86</td>
<td>7.84</td>
<td>0.11</td>
<td>95% RCI CIs</td>
</tr>
</tbody>
</table>

**Notes.** Broglio et al. (2007) Pearson *r* correlations were provided via personal communication (S. P. Broglio, personal communication, February 13, 2012). Schatz (2010) and Elbin et al. (2011) RCI CIs (reliable change index confidence intervals) were obtained using Jacobson and Truax (1991) formula. ICCs = Intraclass correlation coefficients; *VeM* = Verbal memory; *ViM* = Visual memory; PS = Processing speed; RT = Reaction time.

4. SKE proceed to assert that in our consideration of ImPACT’s visual memory domain, we stated that “significant practice effects have been observed for processing speed and visual memory” and then comment “but it is unclear what specific data this assertion is based on.” However, we included the direct references within the same sentence: “significant practice effects in healthy controls have been observed for the processing speed (Iverson, Lovell, et al., 2003) and visual memory (Schatz, 2010) composites” (Mayers & Redick, 2012, p. 237).
5. SKE stated that “one might expect variation in human performance over a 2-year period.” (p. 430). We agree with this statement, especially when considering the still-maturing adolescent period of life. However, the variation described by SKE would seem problematic for using ImPACT with the baseline approach, given the guidelines on the ImPACT website: “We suggest that athletes are tested every two years from 6th grade to senior year of high school. In college, athletes should only be tested once. At the professional level, each athlete should be tested once” (Impact Applications, Inc., 2012b). Cognitive development is still occurring and does not peak until the mid-20s and then declines throughout the 30s for abilities such as Processing Speed (Salthouse, 2009). The current protocol of baseline testing would seem to provide an inaccurate comparison to be used against postconcussion performance multiple years later.

6. We were not clear enough in our statement about significant practice effects observed in ImPACT. We were specifically referring to studies of nonconcussed subjects that found a significant improvement in an ImPACT composite score from Time 1 to Time 2. In our article, we stated that significant practice effects had been observed for Processing Speed (Iverson, Lovell, & Collins, 2003) and Visual Memory (Schatz, 2010). In addition, Elbin et al. (2011) observed significant practice effects for Visual Memory, Processing Speed, and Reaction Time (all ps = .001). In contrast to SKE’s statement, Miller, Adamson, Pink, and Sweet (2007) is another example of a study in which practice effects on ImPACT were observed. Specifically, Miller et al. tested football players at pre-, mid-, and postseason. The effects of practice were seen for Verbal Memory (p = .06), Visual Memory (p = .04), Processing Speed (p = .05), and Reaction Time (p = .04). As we stated in the original article, practice effects complicate the goal of baseline testing, because the comparison with the baseline is affected by the increase in performance due to practice and the decrease in performance assumed to be caused by the concussion.

7. In a later comment, SKE state that “Further, they selectively utilized only one of five statistical techniques presented in the Schatz (2009 [sic]) paper.” (p. 430). In fact, Schatz (2010) reported four techniques: Pearson’s r, ICCs, RCI’s, and RBM. In our paper, we discussed Pearson’s r and ICCs saying “Schatz (2010) reported test–retest intraclass correlations of .45 to .74 and Pearson r test–retest correlations of .30 to .60” (Mayers & Redick, 2012, p. 237).

Sensitivity and specificity

Schatz, Pardini, Lovell, Collins, and Podell (2006) administered ImPACT to 72 concussed high-school athletes and 66 nonconcussed controls. They reported sensitivity to be 82% (nearly 1 in 5 concussed patients incorrectly classified as “normal”) and specificity 89% (1 in 10 nonconcussed athletes incorrectly classified as concussed). Similarly, Van Kampen, Lovell, Pardini, Collins, and Fu (2006) administered ImPACT to 122 concussed high-school and college athletes and 70 nonconcussed controls. Although an abnormal result in a single ImPACT test domain had a sensitivity of 83%, specificity was only 70% (almost 1 in 3 nonconcussed athletes incorrectly classified as concussed). It seems reasonable to conclude that a test to guide “overall clinical management issues” for concussed athletes and their “safe return to participation” should demonstrate more robust sensitivity–specificity characteristics than those shown by these studies, especially when used to guide medical decisions concerning brain-injured patients and their safety after RTP.

Validity issues

1. SKE state that “Finally, it is unclear why the divergent validity reported by Iverson and colleagues (Iverson, Gaetz, Lovell, & Collins, 2005; Iverson et al., 2003) should be ‘interpreted with caution’, or why the authors did not discuss the specific findings from these papers.” (p. 431). We believe that these two citations are in error. We stated “taken with caution” and within the same sentence pointed out that the athletes were still suffering from concussive effects at the time the two cited Iverson studies were conducted. We believe that we described the pertinent findings from the following papers: Iverson, Franzen, Lovell, and Collins (2003); Iverson, Lovell, and Collins (2005). They have addressed the divergent validity issue, but only within the context of athletes still suffering from the acute effects of concussion. We conclude that “given this limitation, the following results should be taken with caution” (Mayers & Redick, 2012, p. 238). We then spent an entire paragraph addressing the results from these two Iverson et al. studies.
2. Later in their critique, SKE say “Mayers and Redick comment on methodological differences [...] which they then use to make broad generalizations regarding ImPACT. [...] However these authors fail to consider that Iverson et al. included a 7–21-day retest interval in their analyses.” (p. 432).

We really don’t know what SKE are trying to say here. We thoroughly discussed this study in our paper, and our “broad generalizations” concerning ImPACT are not simply based upon “methodological differences” but an entire consideration of the literature that we reviewed.

3. SKE assert that we “seemed to lack an understanding of the validity types.” (p. 431). Our discussion of divergent/discriminant validity was not at the level of the entire test but, as we stated, “among the four composites assessed” (p. 238) in ImPACT. They then state that our “discussion of the lack of divergent validity of ImPACT was primarily a discussion of ImPACT’s internal structure and interrelationships, which is reflective of construct validity.” (p. 431). We are sure they know that divergent/discriminant validity is part of construct validity, not separate from it. Later, SKE state that “orthogonality refers to low interrelationships rather than ‘uncorrelated’.” (p. 431). This is incorrect as orthogonal variables are variables that are not correlated. In factor analysis, orthogonal factors have a factor correlation of zero.

More germane to the discussion of ImPACT, and as we discussed in our article, the overlap among the four ImPACT composites is a serious problem for clinical RTP decisions that are based on the number of abnormal ImPACT scores that have yet to return to baseline. Van Kampen et al. (2006) explicitly indicated that their work provided preliminary support for the use of the number of abnormal ImPACT scores as part of the diagnosis, “with a higher number of abnormal composite scores suggesting a more severe concussion” (p. 1634). Again, as we stated in the original article, if the ImPACT composites are correlated, even moderately, than attempting to make decisions based on the number of abnormal composites is misguided.

RTP studies

The time required for complete healing of concussive brain injury is currently unknown. For several decades, athletes were allowed to RTP post concussion when they felt able to participate. Utilizing this management paradigm, it was reported that 15% of National Football League players returned to play “immediately” after a diagnosed concussion, 32% returned during the same game, while more than 90% returned to compete within 7 days (Pellman, Viano, Casson, Arfken, & Feuer, 2005). This model apparently reflected the athletes’ tolerance for unpleasant symptoms and the pressures to RTP, both self-imposed and externally imposed by coaches, peers, and the culture of football. During this period, it also became customary for high-school and college athletes to RTP within 7 days. Not surprisingly, the 7-day interval coincided with the next scheduled football game. Currently, many coaches, athletes, and administrators still expect that a concussed athlete can and should RTP in one week.

Although this practice lacks a scientific basis, accumulated experience has shown that, for many concussed athletes, symptoms resolve by 7 days post injury, and when they RTP they are able to perform their sport adequately. Since these athletes are able to “function” on the field one week following injury, it follows that functional performance tests (i.e., balance, exercise stress, and neuropsychological) probably normalize within a similar interval, and generally they do (McCrea et al., 2003). This is not to say that administering these tests, especially the neuropsychological ones, cannot inform us about the clinical consequences of concussive injury but only to emphasize that a normal result on any performance test (the most commonly utilized one being ImPACT) does not equate with complete healing of a brain injury. In addition, recalling that acutely concussed athletes commonly exhibit attention–concentration–memory deficits and photophobia (sensitivity to light), administering computerized tests of intellectual functions shortly after injury to detect a decrease from “baseline results” violates consensus treatment guidelines that emphasize complete physical and cognitive rest (McCrory et al., 2009) and may in fact delay recovery.

DISCUSSION

Unfortunately, no simple, affordable, and accessible tests to answer the critical question concerning the interval required for healing after a concussive brain injury are yet available. However, complex and expensive tests obtainable mainly in research medical center settings have been developed and utilized to measure sport-related brain injury and to estimate its duration. These techniques include electronic recordings.
of evoked brain wave activity (De Beaumont, Brisson, Lassonde, & Jolicoeur, 2007; Gosselin, Theriault, Leclerc, Montplaisir, & Lassonde, 2006), multitask effects on balance (Halterman et al., 2006; Parker, Osternig, van Donkelaar, & Chou, 2006; van Donkelaar, Osternig, & Chou, 2006), balance testing with destabilized posture (Slobounov, Slobounov, Sebastianelli, Cao, & Newell, 2007), transcranial magnetic spectroscopic measurements of brain metabolites (Henry et al., 2011; Vagnozzi et al., 2008), positron emission tomographic measurements of energy metabolism (Bergsneider et al., 2001), diffusion tensor imaging of injured brain regions (Cubon, Putakian, Boyer, & Dettwiler, 2011), and a recent functional magnetic resonance imaging (fMRI) study (Talavage et al., 2012). Without a lengthy consideration of the complexities of these procedures, the results, summarized in Table 2, indicate that healing of brain damage is not complete within 4 weeks following concussive injury.

Although we and SKE have herein debated interpretations of technical issues related to neuropsychological testing in general and to ImPACT particularly, the major consideration has not been addressed. In scientific inquiry, it is necessary to postulate axioms (fundamental principles) and then to demonstrate how the findings follow logically from these postulates. The ImPACT test’s axioms may be expressed as: uninjured state = baseline score; brain-injured state = significantly reduced score; healed brain injury = return to baseline score. As cited above, ImPACT and other functional tests do not measure recovery from a brain injury, only resolution of some of the manifestations of the injury. Therefore, they cannot accomplish ImPACT’s stated claim that “Our approach to managing concussion has been found to be a reliable and valid method for determining when an athlete has recovered sufficiently from a concussion in order to return to athletic competition” since no one advocates the exposure of an incompletely recovered brain-injured patient to repeated head injury.

The divide between the RTP time interval of one week resulting from typical application of “usual and customary” consensus guidelines that are largely based on symptoms resolution and functional tests (McCrory et al., 2009) and the evidence from the various measurements of brain healing cited above that injury persists for 4 or more weeks is disturbing. Both clinical and experimental evidence have shown that the risk of experiencing a second concussion and the consequences of a second episode are greatest within the two weeks following a first episode (Guskiewicz et al., 2003). These consequences may include rapid uncontrolled brain swelling, resulting in death (McCrory, Davis, & Makdissi, 2012). Although occurring very rarely, the result is catastrophic.

SKE state that we “essentially offer no alternative solution” (p. 430), presumably referring to ImPACT testing, to determine time of RTP. We believe that functional tests do not provide the “solution” for the reasons cited above and that the RTP decision is a strictly clinical one to be decided by a competent, qualified, and experienced practitioner, taking into account the athlete’s age, gender (Dick, 2009), previous head injuries, severity and duration of symptoms and signs, and associated learning disability (Collins et al., 1999). We note that these

### TABLE 2

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Abnormal response duration</th>
<th>Patient group</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEG ERP response</td>
<td>5 weeks</td>
<td>CA, S and A</td>
<td>Gosselin et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>30 months</td>
<td>CA, multiconcussions</td>
<td>DeBeaumont et al. (2007)</td>
</tr>
<tr>
<td>TMS</td>
<td>9 months</td>
<td>CA, multiconcussions</td>
<td>DeBeaumont et al. (2007)</td>
</tr>
<tr>
<td>Multitask-challenged gait stability</td>
<td>28–30 days</td>
<td>CA</td>
<td>Halterman et al. (2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parker et al. (2006)</td>
</tr>
<tr>
<td>Destabilized posture balance testing</td>
<td>14 days</td>
<td>CA, first concussion</td>
<td>Slobounov et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>28 days</td>
<td>CA, multiconcussions</td>
<td>Slobounov et al. (2007)</td>
</tr>
<tr>
<td>PET scan glucose uptake</td>
<td>2–4 weeks</td>
<td>Nonathletes with TBI</td>
<td>Bergsneider et al. (2001)</td>
</tr>
<tr>
<td>DTI MRI</td>
<td>30 days</td>
<td>CA</td>
<td>Cubon et al. (2011)</td>
</tr>
<tr>
<td>fMRI</td>
<td>1–3 months</td>
<td>CA, S and A</td>
<td>Talavage et al. (2012)</td>
</tr>
</tbody>
</table>

Note. CA = concussed athletes; S = symptomatic; A = asymptomatic; EEG = electroencephalography; ERP = event-related potential; TMS = transcranial magnetic stimulation; PET = positron emission tomography; DTI = diffusion tensor imaging; MRI = magnetic resonance imaging; fMRI = functional magnetic resonance imaging; TBI = traumatic brain injury.
significant variables and confounders cannot be factored into the results obtained with any of the functional tests.

We reiterate our conclusion that “for evaluating and advising concussed athletes when to return to play, ImPACT test results should not be the determining factor. Clearly, more research and consideration by neuropsychologists will be necessary to weigh the validity of this conclusion” (Mayers & Redick, 2012, p. 240). Supporting our position is a more recent posting from the ImPACT website stating that “ImPACT assists doctors in making RTP decisions and should never be used as a standalone tool” (Impact Applications, Inc., 2012d). Since ImPACT test results do not relate to time of healing of concussive brain injuries, we fail to see how they assist doctors in making RTP decisions as claimed on their website.

Our suggested “solution” to the current RTP issue is that, in the absence of available technology to precisely determine the time of postconcussion healing of brain damage in individual athletes, all of the methods cited above for determining persistence of brain injury should be utilized to study different injured cohorts (by age, gender, sport, injury history, etc.) to determine mean and standard deviation intervals for healing (resolution of metabolic, physiological, and anatomic abnormalities) of postconcussive brain injury. These results could then be used to provide evidence-based RTP advice to the vastly larger number of the athletes that we serve who are without access to these studies. Until this information becomes available, we advocate for adoption of a more conservative approach to the RTP issue allowing for longer recovery intervals closer to 4–6 weeks than 1 week (Mayers, 2008).

In the future, with new information, it is possible that this interval will need to be extended.

REFERENCES


*Journal of the American Medical Association, 282*(10), 964–970.


