Title: Is it sandbagging or is it just apathy? Baseline ImPACT and external validity indicators in collegiate athletes

Authors: Kate L. Higgins[1,2], Heather C. Bouchard[2,3], Julia E. Maietta[4], Julia M. Laing[2,3], Douglas H. Schultz[2,3]

[1]Department of Athletics, University of Nebraska-Lincoln
[2]Center for Brain, Biology and Behavior, University of Nebraska-Lincoln
[3]Department of Psychology, University of Nebraska-Lincoln
[4]Department of Psychiatry & Behavioral Sciences, University of Oklahoma Health Sciences Center

Corresponding author: Kate Higgins, UNL Athletics, 800 Stadium Drive, Lincoln, NE 68588, United States. Email: khiggins@huskers.com

Abstract

Objective: Baseline cognitive testing is important for sport concussion management. Assessing performance validity during baseline testing is vital to ensure an athlete provides best effort. Stand-alone performance validity tests and embedded validity indicators (collectively called “effort measures” hereafter) are commonly used. This study further investigates efficient ways to assess variable effort in baseline cognitive testing.

Methods: 231 NCAA athletes completed baseline assessment with ImPACT and one of two stand-alone effort measures: Medical Symptom Validity Test or Rey Dot Counting Test. The relationship between effort and cognitive performance, symptoms, and sleep was assessed. **Results:** Athletes who failed two or more effort measures performed worse on the Visual Motor Speed Composite while athletes failing three or more measures performed worse on the Reaction Time Composite. Those athletes who failed effort measures and also reported poor sleep performed worse on both composites. Self-reported symptoms and hours of sleep were not independently associated with failing effort measures. Lastly, athletes with ADHD and/or LD were more likely to fail two+ indicators (46% versus 27% of neurotypical athletes).

Conclusions: Consistent with previous research, ImPACT Baseline++ only identified 1.7% of our sample as poor effort despite 4.8% failing a combination of stand-alone/embedded effort measures and an additional 42.9% failing at least one embedded measure alone. Results suggest that assessing effort on a continuum and using multiple effort measures may be useful to identify suboptimal effort for non-intentional reasons.

Keywords: concussion, performance validity, baseline testing, ImPACT

Public Significance Statement

Baseline ImPACT testing is common and often mandated prior to sport participation, however, valid baselines require athletes to provide their best effort. This study adds to the growing body of literature demonstrating that athletes frequently do not provide best effort on baseline testing. Additionally it suggests that athletes may not be intentionally underperforming, as stand-alone performance validity tests miss a significant number of athletes who have other indicators of poor effort within their ImPACT baselines.

Is it sandbagging or is it just apathy? Baseline ImPACT and external validity indicators in collegiate athletes

Concussion baseline testing with some combination of cognitive, symptoms, and balance/vestibular/oculomotor evaluation has become common and is required at many levels of sport participation. At the NCAA level, the evaluation of these three elements are required for varsity collegiate athletes as a part of their pre-participation evaluation (National Collegiate Athletic Association Sport Science Institute, 2022).

The premise of baseline testing is that an athlete can serve as their own standard for “recovered” should they receive a concussion (Bailey et al., 2009). However, variability in the testing environment or in individual factors, including effort, can impact the validity of baseline testing scores resulting in a potentially inaccurate standard for “normalized” functioning after a concussion. While there is mixed evidence regarding the rate of poor effort in athletes on baseline cognitive testing, a recent systematic review (Gaudet & Weyandt, 2017) found that between 2.7% and 27.9% have indications of poor effort using a variety of validity indicators within Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) a computerized neurocognitive screener commonly used for baseline concussion evaluations in the sport setting (Gaudet & Weyandt, 2017). Poor effort on baseline has serious clinical implications. If an athlete’s data does not provide a valid representation of their individual baseline functioning, they are at risk of being returned to sport and head injury exposure before they are fully recovered, which puts them at increased risk for a subsequent concussion (Bailey et al., 2006). As a result, assessing for sufficient effort in the context of baseline testing has begun to be addressed in the literature, particularly around ImPACT. ImPACT contains several “built-in” embedded performance validity indicators, with associated cut-scores, that are used by the program to identity invalid

baselines (labeled “Baseline++”). As will be discussed further, these do not appear to be particularly sensitive to insufficient effort, missing up to 20% of test-takers who are instructed to intentionally underperform on ImPACT testing (Gaudet & Weyandt, 2017). Because of this, additional embedded indicators of poor effort have been identified in the literature to improve the sensitivity of invalid baseline classification (Higgins et al., 2017; Manderino & Gunstad, 2018; Manderino et al., 2019; Schatz & Glatts, 2013), but have not been incorporated into the ImPACT program.

Assessing for effort - formally called performance validity - is a common and important aspect of traditional, paper-and-pencil neuropsychological assessment. Good effort is characterized by full engagement and best performance on cognitive tasks, while “poor” effort is much more ambiguous and difficult to define. At the extreme of “poor” effort on the effort spectrum is malingering (i.e., intentional poor effort for some external incentive), or in sport concussion contexts, sandbagging (intentionally performing poorly to be cleared for return-to-play status more quickly after concussion). If effort exists on a spectrum, as suggested by Heilbronner et al. (2009), then there is space between those two poles where effort is neither fully “good” nor fully “bad”. In traditional neuropsychological evaluations, it is common to employ a combination of stand-alone performance validity tests (PVTs), a test conducted for the sole purpose of evaluating performance validity/effort,- and embedded performance validity indicators (scores from a test given to measure a different cognitive domain (Heilbronner et al., 2009; Miller et al., 2017; Sweet, Heilbronner, et al., 2021). While both types of effort measures are often employed in neuropsychological testing, there is a breadth of literature demonstrating higher sensitivity and specificity of stand-alone PVTs for intentional poor effort compared to embedded validity indicators (Curtis et al., 2008; Kanser et al., 2019; Ovsiew et al., 2020; Pliskin. Recent consensus guidelines on performance validity assessment suggest the use of stand-alone PVTs as ideal in

clinical practice (Sweet, Heilbronner, et al., 2021). Additionally, recent survey data suggests consensus in the field that stand-alone PVTs are of high importance for neuropsychological evaluations (Schroeder et al., 2016; Sweet, Klipfel, et al., 2021a, 2021b).

However, this recommendation has not changed how baseline concussion assessments are typically conducted in real-world settings. A survey of health care professionals that manage sport-related concussions revealed that it is not common to use either stand-alone PVTs or thorough evaluation of additional embedded validity indicators with ImPACT (Covassin et al., 2009). Gaudet & Weyandt (2017) found, in a systematic review of the literature, that ImPACT Baseline++ validity indicators may miss up to 20% of “fake bad” simulators. Taken in conjunction with a general lack of external PVT assessment use, this data suggests some athletes may have baselines that do not reflect their optimal functioning. More recent literature suggests that between 50% and 60% of simulators remain undetected by ImPACT Baseline++ implying a need to re-evaluate the built-in ImPACT validity indicators (Abeare et al., 2019; Messa et al., 2020; Raab et al., 2020). Other literature has identified additional embedded indicators of invalidity within ImPACT with greater utility than those used by Baseline++. Specifically, Schatz & Glatts (2013) found that the total number of correct distractors for both Word Memory (< 22) and Design Memory (< 16) identified 90 to 100% of simulated sandbagging. Additionally, Higgins et al. (2017) identified a logistic regression equation that demonstrated 100% sensitivity to simulated sandbagging.

Several studies address the use of stand-alone PVTs in addition to ImPACT testing (Abeare et al., 2019; DaCosta et al., 2021; L. Manderino & Gunstad, 2018; Salazar et al., 2017; Schatz & Glatts, 2013). Specifically, Schatz & Glatts (2013) found that the Medical Symptom Validity Test (MSVT) outperformed the built-in ImPACT validity indicators in an experimental sample of college-aged naive and coached

malingerers/sandbaggers. The MSVT identified 80 to 90% of naive and coached malingerers compared to 60 to 75% identified by the built-in ImPACT validity indicators, respectively. Manderino & Gunstad (2018) also employed an experimental malingering/sandbagging paradigm and found that built-in ImPACT validity indicators demonstrated much lower sensitivity than a stand-alone PVT (Word Memory Test) (i.e., the average sensitivity of ImPACT indicators was 0.25 compared to 0.52 for WMT). Two studies have investigated the use of the Dot Counting Test (DCT) in sport concussion baseline testing (DaCosta et al., 2021; Salazar et al., 2017). Salazar et al. (2017) used naive and coached malingering/sandbagging groups to establish the DCT as an acceptable measure of performance invalidity in this population. DaCosta et al. (2021) used a naturalistic sample of collegiate athletes' baselines and found that the DCT was more sensitive in detecting invalid cases than built-in ImPACT validity indicators alone. Their sample also included a small subsample of athletes with attention-deficit/hyperactivity disorder (ADHD) or learning disorders (LD), who were more likely to be flagged by built-in ImPACT validity indicators but did not produce invalid effort on the DCT. DaCosta et al. (2021) is one of the first to compare performance on ImPACT embedded validity indicators to stand-alone PVTs in a population with neurodevelopmental diagnoses. There is additional research which has similarly suggested that athletes with neurodevelopmental disorders may be flagged as invalid on ImPACT more frequently than their neurotypical counterparts (Gaudet & Weyandt, 2017; Maietta, Barchard, et al., 2021; Maietta et al., in press). Theories to explain these findings include athletes with ADHD and/or LD may fall below embedded validity thresholds due to underlying reading and/or

attentional difficulties rather than effort issues. Further research is needed in this regard given DaCosta et al. (2021)'s findings.

The focus on experimental malingering/sandbagging paradigms in all but one study in the extant literature may be problematic, as these patterns of performance may produce exaggerated profiles that are easily detected by traditional PVTs. Both sandbagging and malingering involve the intention to underperform. This intentionality may not be present in all athletes, and may be better explained by the concept of effort existing on a continuum (in contrast to as a dichotomy) initially laid out in the American Academy of Clinical Neuropsychology's (AACN) 2009 consensus statement on validity assessment (Heilbronner et al., 2009) and further discussed in their 2021 follow-up consensus statement (Sweet, Heilbronner, et al., 2021). There is also evidence that many athletes may underperform with more subtle methods than those used by research participants simulating sandbagging or by patients actively attempting to malinger during neuropsychological testing (Abeare et al., 2019). Indifference, dislike of the test due to multiple exposures, or lack of understanding of ImPACT baseline's utility in the management of sport concussion may also play a role in an athlete's engagement in testing or variability in engagement across the course of testing (Erdal, 2012; McClure et al., 2014; Rabinowitz et al., 2015; Schatz et al., 2017; Walton et al., 2018). If the effort supplied by an athlete falls on the middle of the spectrum (rather than on the "good" end or the "poor" end), there may also be more interference in cognitive scores from poor sleep, emotional distress, cognitive fatigue, etc. given there is less engagement to help mitigate the effects of these factors. The concept of "transient validity" is also raised in the AACN 2021 follow-up consensus statement (Sweet, Heilbronner, et al., 2021). This describes failure on a single effort measure in the absence of intentional poor effort. This concern supports the use of multiple validity checks (both embedded

and stand-alone) throughout a neuropsychological evaluation. While baseline concussion testing is significantly shorter than a full neuropsychological evaluation, the concept of multiple validity checks may be useful in differentiating less intentional "middle of the effort spectrum" performance from intentional underperforming. Neurodevelopmental diagnoses, such as ADHD or LD, can make the classification of effort even more variable (Manderino et al., 2019). These factors individually or in combination could easily play a role in athletes providing less-than-best effort without intentionally providing poor effort or sandbagging.

Therefore, there is a need to further investigate efficient ways to evaluate effort in baseline cognitive concussion assessment- either through stand-alone PVTs or embedded effort indices on ImPACT. Additionally, further evaluation on the agreement (or lack of agreement) between embedded ImPACT invalidity indicators and stand-alone "gold-standard" PVT measures in identifying poor effort may also be useful. Given the well documented utility of stand-alone PVTs in detecting intentional poor effort in traditional neuropsychological batteries, adding a stand-alone PVT to supplement baseline ImPACT in the detection of invalid performance is consistent with best practice guidelines in the field (Abeare et al., 2019; DaCosta et al., 2021; Sweet, Heilbronner, et al., 2021), but questions remain about their utility in detecting less-than-best effort that may be related to factors other than intentional poor effort and the return-on-investment of adding an additional measure to concussion baselines.

Methods

Participants

NCAA Division I collegiate athletes completed baseline assessments annually (with two years of data examined for this research) as part of their standard clinical pre-participation evaluations. All athletes completed ImPACT (N = 231). In addition to ImPACT, athlete completed the Medical Symptom Validity Test (n = 130) in year one,

and athletes completed the Rey Dot Counting Test (*n* = 101) in year two. Ages ranged from 17 to 24 (*M* = 18.6*, SD* = 1.3; 63% male). Race and ethnicity were self-reported by participants, and at the time of data collection, were collected as a single variable instead of two. Participants self-identified as White (67%), African American/Black (17%), Biracial/Multiracial (10%), Latinx/Hispanic (4%), and Asian (2%). The majority of participants were native to the United States (92%). A subset of participants self-reported diagnoses of ADHD, LD, or a combination of the two - hereafter designated ADHD and/or LD (*n* = 44). Participants were from 15 sports, including baseball, basketball, bowling, football, golf, gymnastics, rifle, soccer, softball, swimming, tennis, track & field, volleyball, wrestling, and cross-country.

Materials

ImPACT

ImPACT is a computerized neuropsychological assessment used to assess cognitive performance at baseline and recovery following concussion by comparing post-concussion scores to baseline performance (ImPACT Applications, 2021). Four composite scores (Verbal Memory, Visual Memory, Visual-Motor Speed, and Reaction Time) are calculated from 10 cognitive tasks. In addition to cognitive assessment, athletes self-report their total symptoms, and optionally report their total hours of sleep the night prior to testing. Embedded performance validity indicators built into ImPACT (Baseline++) are: Impulse Control Composite > 30, Word Memory Learning Percent Correct < 69%, Design Memory Learning Percent Correct < 50%, Three Letters Total Letters Correct < 8 (ImPACT Applications, 2021) and X's and O's Total Incorrect > 30 (ImPACT Applications, 2016). Additional embedded indicators (that are not built into ImPACT) include: Word Memory Correct Distractors (WMCD)< 22 and Design Memory

Correct Distractors (DMCD) < 16 (Schatz & Glatts, 2013) and a logistic regression equation (LRE) using Word Memory Learning Percent Correct, Word Memory Delay Memory Percent Correct, Design Memory Total Percent Correct, and X's and O's Total Correct (Interference) (Higgins et al., 2017). When evaluating cognitive tasks, only the full composites for Visual Motor Speed and Reaction Time were evaluated, as portions of both Verbal Memory and Visual Memory composites are used in the embedded effort measures described above. Individual variables from each of Verbal Memory and Visual Memory composites were evaluated as a proxy (Three Letters-Total Sequence Correct for Verbal Memory composite and X's and O's- Total Correct/Memory for Visual Memory composite) as these variables are part of the composites, but are not included in any of the embedded effort measures assessed in the following analysis.

Rey Dot Counting Test (DCT)

The Rey Dot Counting Test (DCT) is a timed task designed by Rey (1941) to detect poor effort with a total administration time of less than 5 minutes. The test consists of twelve index cards containing differing numbers of dots. Several scores are calculated from the test: mean counting times, ratio (mean time to count different groups of dots), errors (total number of errors), and a combination score (mean dot counting time for each group + number of errors) (Boone et al., 2002). A cut-score of ≥15 was used to identify poor effort (DaCosta et al., 2021). ***Medical Symptom Validity Test (MSVT)***

The MSVT (Green, 2004) is a computerized, stand-alone PVT. The MSVT takes approximately 15-20 minutes to complete, including a 10- minute delay and delayed memory trials. A list of related word pairs are presented on a computer screen and then the participant has to identify the learned words using a variety of memory tasks. Immediate Recall, Delayed

Recall, and Consistency variables were analyzed with ≤85% on any of the three variables signifying poor effort (Green, 2004).

Procedure

As part of their standard clinical pre-participation concussion evaluation, new student-athletes completed a series of brief assessments, including PVTs, administered by either a sports neuropsychologist or trained doctoral students. ImPACT test administration occurred in a group setting using a standardized administration protocol. PVT's were completed in an individualized setting. ImPACT and the PVT included in baseline assessment for that year (either DCT or MSVT) were all administered on the same day, but possibly not back-to-back (due to administration order of the multiple assessments completed as a part of concussion baseline evaluation). The use of student athletes' deidentified data for research purposes, and the use of these data for this specific study, was approved by the home institution's Institutional Review Board (IRB).

Statistical Analysis

Statistical analyses were conducted in R (RStudio Team, 2021). The relationship between effort and clinical outcomes, such as cognitive tasks, total symptom score, and sleep was assessed using linear models with the *lm* command. Effort was evaluated across a continuum (zero, one, two and three+ failed effort indicators) within the linear models, except in the interaction models. In these models effort was evaluated as a binary variable (zero or one+ failed effort indicators). Linear models were also used to assess the relationship between ADHD and/or LD diagnosis and clinical outcomes. The relationship between categorical variables, such as effort and ADHD and/or LD diagnosis, was assessed using a chi-square test of independence with the *chisq.test* commands.

Results

Records for 231 collegiate athletes' clinical concussion baseline assessments were evaluated. Of the 231 athletes, 110 failed at least one sign of poor effort (either on an embedded indicator, a PVT, or a combination of the two). Thirty-nine athletes failed only one effort indicator, 40 athletes failed two effort indicators, and 31 athletes failed three or four (three+) effort indicators. Table 1 displays these results and the number of athletes who failed each individual effort indicator. Neither self-reported total symptoms ($F(3, 227) = 0.92$, $p = 0.431$) nor self-reported hours of sleep the previous night ($F(3, 184) = 1.18$, $p = 0.318$) were significantly associated with failing one, two, or three+ effort indicators.

Table 1. Number of failed effort indicators across individual effort measures

Effort Measure	**Number (%) identified by each validity indicator**
Dot Counting Test (≥15)	10/101 (9.9%)
Medical Symptom Validity Test (IR, DR, CNS ≤ 85)	1/116 (<1%)
Baseline ++	4/231 (1.7%)
Word Memory Correct Distractors (WMCD) < 22 (Schatz and Glatts, 2013)	60/231 (26.0%)
Design Memory Correct Distractors (DMCD) < 16 (Schatz and Glatts, 2013)	68/231 (29.4%)
Logistic Regression Equation (LRE) ≥ 0.23 (Higgins et al., 2017)	72/231 (31.2%)
Fail only stand alone PVT only (DCT or MSVT)	4/110 (3.6%)
Fail only embedded effort indicator (Baseline ++, WCMD, DMCD, LRE)	99/110 (90%)
Fail a combination of embedded and stand-alone	7/110 (6.3%)

As expected, the number of failed effort measures was related to performance on cognitive scores. Visual-Motor Speed composite was associated with failed effort measures ($F(3, 227) = 9.90$, $p < 0.0001$) . Figure 1A displays the relationship between

Visual-Motor Speed composite performance and number of failed effort indicators. Visual-Motor Speed performance was significantly worse for athletes who failed two (t = -3.179, p = 0.002) or three+ (t = -4.866, p < 0.0001) effort measures compared to those who did not fail any. Visual- Motor Speed composite was not different for those who failed only one effort measure compared to those who did not fail any (t = -0.369, p = 0.712). Hours of sleep reported also played a role in Visual-Motor Speed composite scores. Figure 1B displays the interaction between binary effort (i.e., zero failed effort indicators and one+ failed effort indicators) and self-reported sleep the previous in relation to Visual-Motor Speed performance. Athletes' Visual-Motor Speed performance was slower when they both failed effort measures and got less sleep compared to athletes who also reported less sleep, but did not fail any effort measures (t = 2.117, p = 0.036).

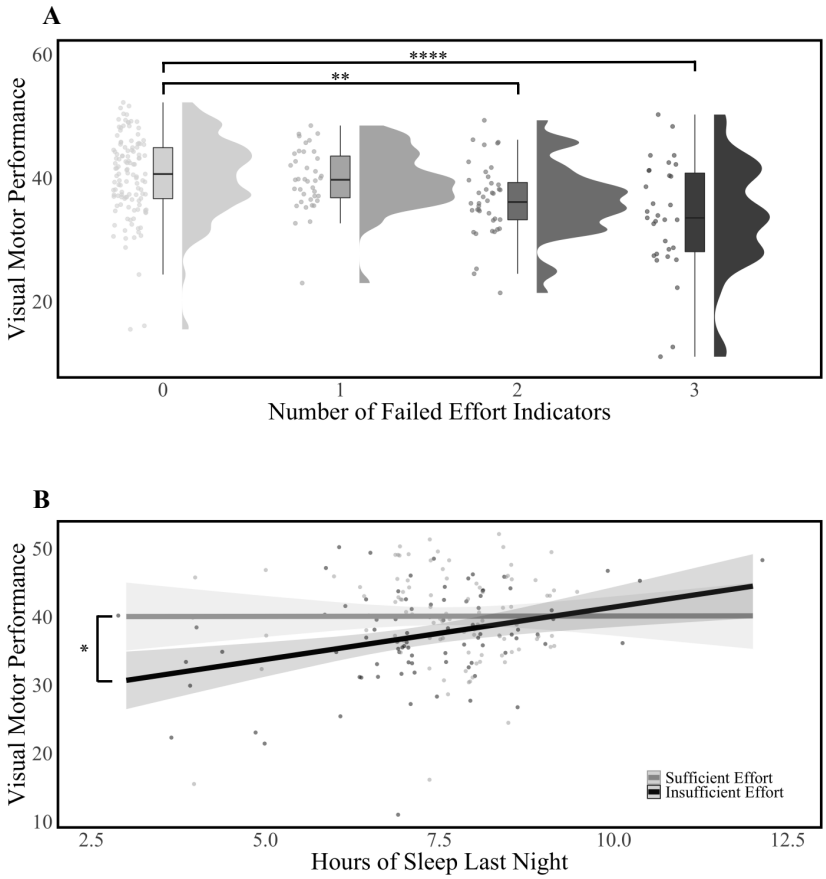

Figure 1. Relationship between effort and Visual-Motor Speed composite Performance on the Visual-Motor Speed composite was associated with the number of failed effort indicators.
A) Athletes who failed two or three+ effort indicators performed significantly worse compared to athletes who did not fail any.
B) Athletes who failed one+ effort indicators performed more slowly on Visual-Motor Speed tasks when they reported less sleep compared to athletes who did not fail any effort indicators.

Number of failed effort indicators was also associated with Reaction Time composite ($F(3, 227) = 5.46$, $p = 0.001$). Figure 2A displays the relationship between Reaction Time composite performance and number of failed effort indicators. Reaction Time composite performance was worse for athletes who failed three+ ($t = 4.038$, $p < 0.0001$) effort indicators compared to those who did not fail any effort indicators. Yet, failing one ($t = 0.990$, $p = 0.323$) or two ($t = 1.141$, $p = 0.255$) effort indicators did not impact performance on the Reaction Time composite compared to those who did not fail any effort indicators. In addition, Figure 2B displays the interaction between binary effort and self-reported sleep in relation to Reaction Time performance. Reaction Time performance was slower in athletes how both reported less sleep and failed one+ effort measure ($t = -2.730$, $p = 0.007$) compared to failing zero effort measures.

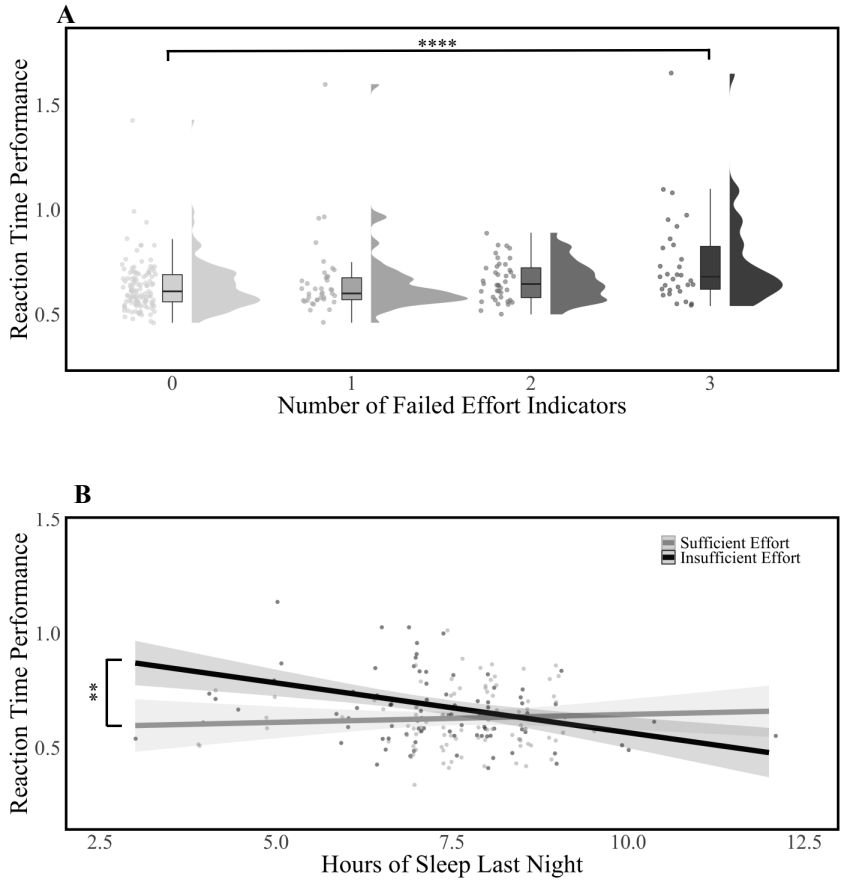

Figure 2. Relationship between effort and Reaction Time composite Performance on the Reaction Time composite was associated with the number of failed effort indicators.
A) Athletes who failed three+ effort indicators performed more slowly compared to athletes who did not fail any effort measures.
B) Athletes who failed one+ effort indicators demonstrated slower scores when they reported less sleep compared to athletes who did not fail any effort indicators.

An individual subtest score was used as a proxy for both Visual Memory and Verbal Memory composites. The number of failed effort indicators was significantly associated with the proxies for Verbal Memory ($F(3, 227) = 4.66$, $p = 0.004$) and Visual Memory ($F(3, 227) = 4.95$, $p = 0.002$) performance. Scores on the Verbal Memory composite proxy were significantly worse for athletes who failed two ($t = -2.483$, $p = 0.014$) or three+ ($t = -2.912$, $p = 0.004$) effort measures compared to those who did not fail any, while failing only one effort measure was not significantly different ($t = 0.335$, $p = 0.752$). Performance on the Visual Memory composite proxy was also significantly worse for athletes who failed two ($t = -2.253$, $p = 0.025$) or three+ ($t = -3.541$, $p < 0.001$) effort measures compared to those who did not fail any, while failing only one effort measure was not significantly different ($t = -0.797$, $p = 0.426$).

ADHD and/or LD diagnostic status also played a role in ImPACT outcomes. Figure 3 displays the relationship between ADHD and/or LD diagnosis with self-reported symptoms, self-reported hours of sleep, and Visual-Motor Speed performance. Athletes who self-reported a diagnosis of ADHD and/or LD reported more symptoms ($t = 4.071$, $p < 0.0001$), less sleep ($t = -2.931$, $p = 0.004$), and performed worse on the Visual-Motor Speed composite ($t = -1.989$, $p = 0.048$). However, athletes with an ADHD and/or LD diagnosis did not perform significantly worse on the Reaction Time composite ($t = 1.622$, $p = 0.106$) compared to athletes who did not endorse a diagnosis.

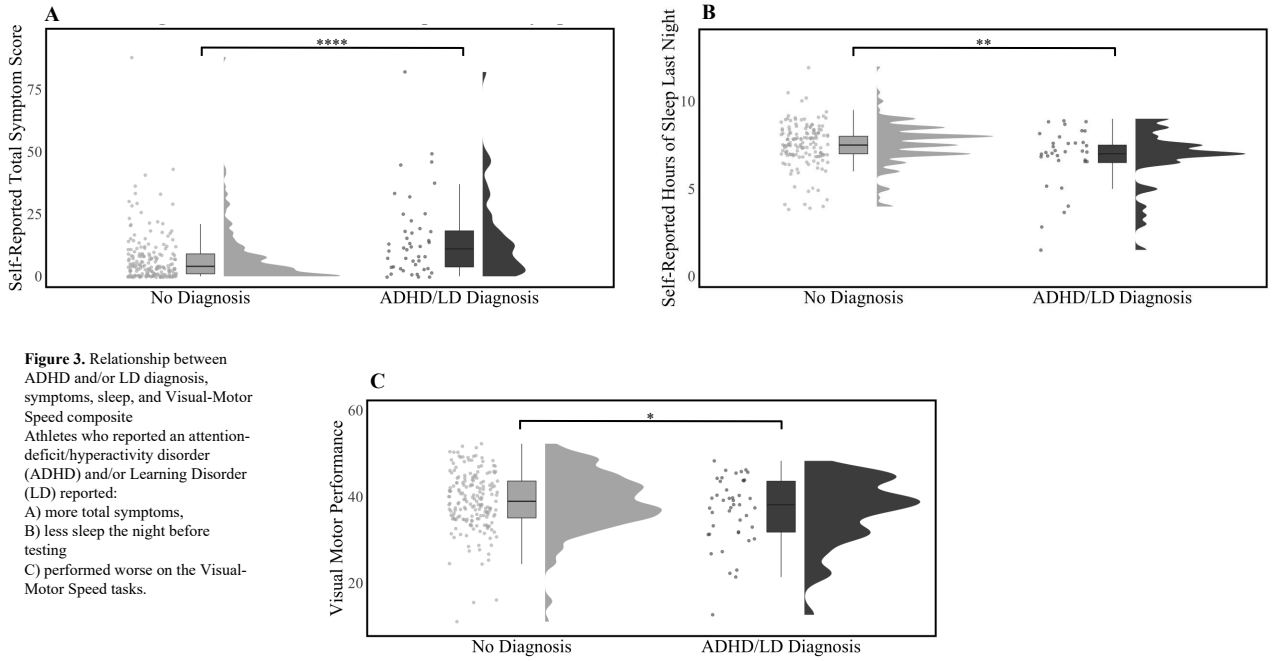

Figure 3. Relationship between ADHD and/or LD diagnosis, symptoms, sleep, and Visual-Motor Speed composite
Athletes who reported an attention-deficit/hyperactivity disorder (ADHD) and/or Learning Disorder (LD) reported:
A) more total symptoms,
B) less sleep the night before testing
C) performed worse on the Visual-Motor Speed tasks.

Table 2 reports information regarding athlete effort and ADHD and/or LD diagnosis. The proportion of athletes who failed one or more effort indicators was related to a diagnosis of ADHD and/or LD (X^2 (1, N = 231) = 4.83, p = 0.028) and athletes who self-reported a diagnosis of ADHD and/or LD were more likely to fail one+ effort measures. A preliminary analysis of athletes diagnosed with ADHD and/or LD (n = 44), assessed effort across a continuum (zero, one, and two+ failed effort measures). The proportion of athletes diagnosed with ADHD and/or LD significantly differed by effort group (X^2 (2, N = 231) = 6.55, p = 0.038). The proportion of athletes failing only one effort measure was similar (18.2% of athletes diagnosed with ADHD and/or LD vs 16.6% of athletes without a diagnosis). In contrast, 45.5% of athletes who were diagnosed with ADHD and/or LD failed two or more effort indicators, while only 27.3% of athletes without an ADHD and/or LD failed two or more effort indicators.

Table 2. Failed effort indicators and ADHD and/or LD diagnosis- **N (%)**

Number of Failed PVTs	**No diagnosis**	**ADHD and/or LD diagnosis**
Zero	105/187 (56.1%)	16/44 (36.4%)
One+	82/187 (43.9%)	28/44 (63.6%)
Effort Across a Continuum		
Zero	105/187 (56.1%)	16/44 (36.4%)
One	31/187 (16.6%)	8/44 (18.2%)
Two	51/187 (27.3%%)	20/44 (45.5%)
Stand-alone PVTs vs Embedded Effort Indicators		
Failed Stand-alone	7/187 (3.7%)	4/44 (9.1%)
Failed Embedded	79/187 (41.7%)	27/44 (61.4%)

Discussion

Identifying less-than-best effort in the context of baseline concussion testing continues to be challenging, but the implications of missing insufficient effort is a problem that cannot be ignored. Consistent with previous research (Gaudet & Weyandt, 2017), the embedded Baseline++ validity indicators built into ImPACT appear to do a poor job of identifying poor effort and identified only 1.7% of our sample as providing poor effort despite 4.8% of the sample failing a combination of stand-alone PVT and embedded validity indicator and an additional 42.9% failing at least one embedded validity indicator. In contrast to previous research on concussion baselines and neuropsychological testing in general (Boone et al., 2002; DaCosta et al., 2021; Green, 2004; Schatz & Glatts, 2013), the prevalence of poor effort identified in our sample by stand-alone PVTs was lower than expected . Even more ironically, but consistent with DaCosta et al. (2021), there was no overlap between those athletes identified by Baseline++ and PVTs.

Also consistent with previous research, the embedded validity indicators suggested by Higgins et al. (2017) and Schatz & Glatts (2013) identify a notably larger number of athletes with questionable effort than Baseline++ or stand-alone PVTs. These discrepancies raise several possibilities. First, it provides support for the theoretically logical, but difficult to operationalize, idea that effort is on a continuum where good effort is on one end, poor effort is on the other, and variable, disengaged, or apathetic (e.g., less-than-best) effort is in the middle. These results may also suggest that less-than-best effort looks different depending on the motivation (or maybe lack thereof), with malingering (intentional poor performance with external incentive) and sandbagging (intentional poor performance on baseline in order to look recovered more quickly after concussion) demonstrating more egregious differences on test scores including effort indices and representing a more extreme end of the continuum. In contrast, the less-than-best effort seen in this study, and likely in other baseline situations, may have less intention behind it- reflecting the middle of the effort continuum, potentially more influenced by apathy and lack of understanding about the utility of baseline testing, or variable engagement rather than any effortful intention to underperform.

These findings would also support the idea that using multiple validity checks across an evaluation is important and a single failed validity measure does not necessarily indicate intentional poor effort. ImPACT composites were significantly different based on the number of validity indicators that an athlete was flagged on, but differences were not seen until the athlete flagged on multiple indicators (2+ for Visual-Motor Speed composite, Verbal Memory composite proxy, Visual Memory composite proxy, and 3+ for Reaction Time composite).

Also similar to previous research (Cook et al., 2017; Gaudet & Weyandt, 2017; Maietta et al., 2021; Maietta et al., in press; Manderino & Gunstad, 2018; Manderino et

al., 2019; Nelson et al., 2016; Schatz et al., 2014; Szabo et al., 2013; Tsushima et al., 2019), ADHD and/or LD diagnoses increased the number of validity indicators flagged. These athletes also reported more symptoms (and less sleep), and demonstrated lower scores on cognitive composites. The ongoing challenge in studying effort in athletes in ADHD and/or LD is differentiating change due to effort from change due to diagnosis (e.g., variable and inefficient attention or reading difficulties). Future research is certainly needed in this regard.

Limitations

There were several limitations to this study. First ImPACT was completed in a group setting while the stand-alone PVTs were administered in an individual setting. Research has shown that administration setting can make a difference in ImPACT composite scores, so this may play a role as athletes may have put forth more effort on PVT's as they were one-on-one with an administrator. Reliance on self-reported neurodevelopmental disorder and hours of sleep the night before testing are also limitations as compared to more objective data such as clinician confirmed diagnosis. Finally, identifying effort in the middle of the effort continuum is significantly more challenging than effort on the ends of continuum, as there are neither clear operational definitions nor tasks or measures to quantify it. Finally, our ADHD and/or LD sample was relatively small, and the subset of the ADHD and/or LD that also failed multiple validity measures was even smaller. This may limit the generalizability of these findings.

Future Directions

Two primary areas for future study were identified by this research. The first is continuing to study effort along a continuum, rather than using the typical dichotomous malingering/sandbagging paradigm. The second is improving understanding of what less-than-best effort looks like in neurodiverse athletes.

Finally, this article adds another data point in the growing body of literature that

questions the utility (both in terms of reliability and return-on-investment) of baseline cognitive testing for managing sport concussion return-to-play decisions. First, there should be an emphasis on assessing effort across a continuum during baseline assessment. Additionally, the answer may not be to do away with baseline testing, but to change the paradigms that we use to evaluate baseline functioning and effort, such as increasing an emphasis on consistency across tasks or trials, creating less time-consuming, but more frequently administered baseline evaluations, or coming up with new baseline cognitive measures that have more ecological validity or utility across disciplines (academics, sport performance psychology, concussion management, etc).

References

Abeare, C., Messa, I., Whitfield, C., Zuccato, B., Casey, J., Rykulski, N., & Erdodi, L. (2019). Performance validity in collegiate football athletes at baseline neurocognitive testing. The Journal of Head Trauma Rehabilitation, 34(4), E20. https://doi.org/10.1097/HTR.0000000000000451

Bailey, C. M., Barth, J. T., & Bender, S. D. (2009). SLAM on the stand: How the sports-related concussion literature can inform the expert witness. The Journal of Head Trauma Rehabilitation, 24(2), 123–130. https://doi.org/10.1097/HTR.0b013e31819c1caa

Bailey, C. M., Echemendia, R. J., & Arnett, P. a. (2006). The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. Journal of the International Neuropsychological Society : JINS, 12, 475–484. https://doi.org/10.1017/S1355617706060619

Boone, K. B., Lu, P., Back, C., King, C., Lee, A., Philpott, L., Shamieh, E., & Warner-Chacon, K. (2002). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. Archives of Clinical Neuropsychology, 17(7), 625–642. https://doi.org/10.1016/S0887-6177(01)00166-4

Cook, N. E., Huang, D. S., Silverberg, N. D., Brooks, B. L., Maxwell, B., Zafonte, R., Berkner, P. D., & Iverson, G. L. (2017). Baseline cognitive test performance and concussion-like symptoms among adolescent athletes with ADHD: Examining differences based on medication use. The Clinical Neuropsychologist, 31(8), 1341–1352. https://doi.org/10.1080/13854046.2017.1317031

Covassin, T., Elbin, R. J., Stiller-Ostrowski, J. L., & Kontos, A. P. (2009). Immediate post-concussion assessment and cognitive testing (ImPACT) practices of sports medicine professionals. Journal of Athletic Training. https://doi.org/10.4085/1062-6050-44.6.639 Curtis, K. L., Thompson, L. K., Greve, K. W., & Bianchini, K. J. (2008). Verbal fluency indicators of malingering in traumatic brain injury: Classification accuracy in known groups. The Clinical Neuropsychologist, 22(5), 930–945. https://doi.org/10.1080/13854040701563591

DaCosta, A., Webbe, F., & LoGalbo, A. (2021). The Rey Dot Counting Test as a Tool for Detecting Suboptimal Performance in Athlete Baseline Testing. Archives of Clinical Neuropsychology, 36(3), 414–423. https://doi.org/10.1093/arclin/acaa052

Erdal, K. (2012). Neuropsychological testing for sports-related concussion: How athletes can sandbag their baseline testing without detection. Archives of Clinical Neuropsychology, 27(5), 473–479. https://doi.org/10.1093/arclin/acs050

Gaudet, C. E., & Weyandt, L. L. (2017). Immediate post-concussion and cognitive testing (ImPACT): A systematic review of the prevalence and assessment of invalid performance. The Clinical Neuropsychologist, 31(1), 43–58. https://doi.org/10.1080/13854046.2016.1220622

Green, P. (2004). Green's Medical Symptom Validity Test (MSVT) User's Manual. Edmonton, Canada: Green's Publishing, Inc.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Participants1, C. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. The Clinical Neuropsychologist, 23(7), 1093–1129. https://doi.org/10.1080/13854040903155063

Higgins, K. L., Denney, R. L., Maerlender, A., & Higgins, K. L. (2017). Sandbagging on the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) in a high school athlete population. Archives of Clinical Neuropsychology, 32, 259–266. https://doi.org/10.1093/arclin/acw108

ImPACT Applications, Inc. (2016). ImPACT: Administration and Interpretation Manual. Retrieved from ImPACT Applications, Inc website: https://impacttest.com

ImPACT Applications, Inc. (2021). ImPACT Version 4: Administration and Interpretation Manual. Retrieved from ImPACT Applications, Inc website: https://impacttest.com/wp-content/uploads/LBL-01_v14_ImPACT-Version-4_Administra tion_Manual.pdf

Kanser, R. J., Rapport, L. J., Bashem, J. R., & Hanks, R. A. (2019). Detecting malingering in traumatic brain injury: Combining response time with performance validity test accuracy. The Clinical Neuropsychologist, 33(1), 90–107. https://doi.org/10.1080/13854046.2018.1440006

Maietta, J. E., Barchard, K. A., Kuwabara, H. C., Donohue, B. D., Ross, S. R., Kinsora, T. F., & Allen, D. N. (2021). Influence of special education, ADHD, autism, and learning disorders on ImPACT validity scores in high school athletes. Journal of the International Neuropsychological Society. https://doi.org/10.1017/s1355617720001149

Manderino, L., & Gunstad, J. (2018). Collegiate student athletes with history of ADHD or academic difficulties are more likely to produce an invalid protocol on baseline impact testing. Clinical Journal of Sport Medicine, 28(2), 111–116. https://doi.org/10.1097/JSM.0000000000000433

Manderino, L. M., & Gunstad, J. (2018). Performance of the immediate post-concussion assessment and cognitive testing protocol validity indices. Archives of Clinical Neuropsychology, 33(5), 596–605. https://doi.org/10.1093/arclin/acx102

Manderino, L. M., Zachman, A. M., & Gunstad, J. (2019). Novel ImPACT validity indices in collegiate student-athletes with and without histories of ADHD or academic difficulties. Clinical Neuropsychologist, 33(8), 1455–1466. https://doi.org/10.1080/13854046.2018.1539191

McClure, D. J., Zuckerman, S. L., Kutscher, S. J., Gregory, A. J., & Solomon, G. S. (2014). Baseline neurocognitive testing in sports-related concussions: The importance of a prior night's sleep. The American Journal of Sports Medicine, 42(2), 472–478. https://doi.org/10.1177/0363546513510389

Messa, I., Korcsog, K., & Abeare, C. (2020). An updated review of the prevalence of invalid performance on the Immediate Post-Concussion and Cognitive Testing (ImPACT). The Clinical Neuropsychologist, 1–24. https://doi.org/10.1080/13854046.2020.1866676

Miller, J. B., Axelrod, B. N., Schutte, C., & Davis, J. J. (2017). Symptom and performance validity assessment in forensic neuropsychology. In APA handbook of forensic neuropsychology (pp. 67–109). American Psychological Association. https://doi.org/10.1037/0000032-005

National Collegiate Athletic Association Sport Science Institute. (2022). Concussion Safety Protocol Checklist. Retrieved from NCAA Sport Science Institute website: https://ncaaorg.s3.amazonaws.com/ssi/concussion/SSI_ConcussionSafetyProtocolChec kli st.pdf

Nelson, L. D., Guskiewicz, K. M., Marshall, S. W., Hammeke, T., Barr, W., Randolph, C., & McCrea, M. A. (2016). Multiple self-reported concussions are more prevalent in athletes with ADHD and learning disability. Clinical Journal of Sport Medicine, 26(2), 120–127. https://doi.org/10.1097/JSM.0000000000000207

Ovsiew, G. P., Resch, Z. J., Nayar, K., Williams, C. P., & Soble, J. R. (2020). Not so fast! Limitations of processing speed and working memory indices as embedded performance validity tests in a mixed neuropsychiatric sample. Journal of Clinical and Experimental Neuropsychology, 42(5), 473–484. https://doi.org/10.1080/13803395.2020.1758635

Pliskin, J. I., DeDios Stern, S., Resch, Z. J., Saladino, K. F., Ovsiew, G. P., Carter, D. A., & Soble, J. R. (2021). Comparing the psychometric properties of eight embedded performance validity tests in the Rey Auditory Verbal Learning Test, Wechsler Memory Scale Logical Memory, and Brief Visuospatial Memory Test-Revised recognition trials for detecting invalid neuropsychological test performance. Assessment, 28(8), 1871–1881. https://doi.org/10.1177/1073191120929093

Raab, C. A., Peak, A. S., & Knoderer, C. (2020). Half of purposeful baseline sandbaggers undetected by ImPACT's embedded invalidity indicators. Archives of Clinical Neuropsychology, 35(3), 283–290. https://doi.org/10.1093/arclin/acz001

Rabinowitz, A. R., Merritt, V. C., & Arnett, P. A. (2015). The return-to-play incentive and the effect of motivation on neuropsychological test-performance: Implications for baseline concussion testing. Developmental Neuropsychology, 40(1), 29–33. https://doi.org/10.1080/87565641.2014.1001066

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.). = The psychological examination in cases of traumatic encepholopathy. Archives De Psychologie, 28, 215–285.

RStudio Team. (2021). RStudio: Integrated Development for R. (1.4.1106) [RStudio]. PBC. http://www.rstudio.com/.

Salazar, C. L., Mulligan, K., & Webbe, F. (2017). Updating the Rey Dot Counting Test for use in collegiate baseline concussion testing. Developmental Neuropsychology, 42(2), 83–92. https://doi.org/10.1080/87565641.2017.1303701

Schatz, P., Elbin, R., Anderson, M., Savage, J., & Covassin, T. (2017). Exploring sandbagging behaviors, effort, and perceived utility of the ImPACT baseline assessment in college athletes. Sport, Exercise, and Performance Psychology, 6(3), 243–251.

https://doi.org/10.1037/spy0000100

Schatz, P., & Glatts, C. (2013). “Sandbagging” baseline test performance on ImPACT, without detection, is more difficult than it appears. Archives of Clinical Neuropsychology, 28(3), 236–244. https://doi.org/10.1093/arclin/act009

Schatz, P., Kelley, T., Ott, S. D., Solomon, G. S., Elbin, R. J., Higgins, K., & Moser, R. S. (2014). Utility of repeated assessment after invalid baseline neurocognitive test performance. Journal of Athletic Training, 49(5), 659–664. https://doi.org/10.4085/1062-6050-49.3.37

Schroeder, R. W., Martin, P. K., & Odland, A. P. (2016). Expert beliefs and practices regarding neuropsychological validity testing. The Clinical Neuropsychologist, 30(4), 515–535. https://doi.org/10.1080/13854046.2016.1177118

Soble, J. R., Webber, T. A., & Bailey, K. C. (2021). An overview of common stand-alone and embedded PVTs for the practicing clinician: Cutoffs, classification accuracy, and administration times. In R. W. Schroeder & P. K. Martin (Eds.), Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance (pp. 126–149). Guilford Press.

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference Participants. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. The Clinical Neuropsychologist, 1–54. https://doi.org/10.1080/13854046.2021.1896036

Sweet, J. J., Klipfel, K. M., Nelson, N. W., & Moberg, P. J. (2021a). Professional practices, beliefs, and incomes of U.S. neuropsychologists: The AACN, NAN, SCN 2020 practice and “salary survey.” The Clinical Neuropsychologist, 35(1), 7–80. https://doi.org/10.1080/13854046.2020.1849803

Sweet, J. J., Klipfel, K. M., Nelson, N. W., & Moberg, P. J. (2021b). Professional practices, beliefs, and incomes of postdoctoral trainees: The AACN, NAN, SCN 2020 practice and “salary survey.” Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 36(1), 1–16. https://doi.org/10.1093/arclin/acaa116

Szabo, A. J., Alosco, M. L., Fedor, A., & Gunstad, J. (2013). Invalid performance and the ImPACT in national collegiate athletic association division I football players. Journal of Athletic Training, 48(6), 851–855. https://doi.org/10.4085/1062-6050-48.6.20

Tsushima, W. T., Yamamoto, M. H., Ahn, H. J., Siu, A. M., Choi, S. Y., & Murata, N. M. (2019). Invalid baseline testing with ImPACT: Does sandbagging occur with high school athletes? Applied Neuropsychology. Child, 1–10. https://doi.org/10.1080/21622965.2019.1642202

Walton, S. R., Broshek, D. K., Freeman, J. R., Cullum, C. M., & Resch, J. E. (2018). Valid but invalid: Suboptimal ImPACT baseline performance in university athletes. Medicine and Science in Sports and Exercise, 50(7), 1377–1384.

https://doi.org/10.1249/MSS.0000000000001592

Whiteside, D. M., Kogan, J., Wardin, L., Phillips, D., Franzwa, M. G., Rice, L., Basso, M., & Roper, B. (2015). Language-based embedded performance validity measures in traumatic brain injury. Journal of Clinical and Experimental Neuropsychology, 37(2), 220–227. https://doi.org/10.1080/13803395.2014.1002758

Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort indicators within the California Verbal Learning Test-II (CVLT-II). The Clinical Neuropsychologist, 24(1), 153–168. https://doi.org/10.1080/13854040903107791