OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 27 (2012) 473–479

Archives of CLINICAL NEUROPSYCHOLOGY

Neuropsychological Testing for Sports-related Concussion: How Athletes Can Sandbag their Baseline Testing Without Detection

Kristi Erdal*

Department of Psychology, Colorado College, Colorado Springs, CO, USA

*Corresponding author at: Department of Psychology, Colorado College, 14 East Cache La Poudre Street, Colorado Springs, CO 80903, USA. Tel.: +1-719-389-6598; fax: +1-719-389-6284. *E-mail address*: kerdal@coloradocollege.edu (K. Erdal).

Accepted 28 April 2012

Abstract

Neuropsychological baseline testing is commonplace in the assessment of concussion; however, claims of "sandbagging" the baseline have led neuropsychologists to ask to what extent athletes can perform intentionally poorly on baseline testing without reaching threshold on the test validity indicators. Seventy-five undergraduate athletes were re-administered the ImPACT neurocognitive battery, which they had previously taken to establish baseline functioning, but were instructed to perform more poorly than their baseline without reaching threshold on the test validity indicators. Eight participants were able to successfully fake significantly lower scores without detection by validity indicators. Concussion history was not related to performance. Successful fakers did not perform significantly worse on the Reaction Time Composite and Three Letters Total Letters Correct, questioning the utility of these measures for detecting "sandbagging." Successful fakers reported using less purposeful faking strategies which naturally facilitated errors. The data suggest that "sandbagging" the baseline, even under conditions involving motivation, instruction, and experience with the test, is difficult to accomplish without being detected.

Keywords: Neuropsychological testing; Athletes; Baseline; Concussion

Introduction

Neuropsychological testing of athletes to manage concussion has long included pre-season or baseline testing as well as post-concussion testing, as needed (Kaminski, Groff, & Glutting, 2009; Solomon & Haase, 2008). The use of post-concussion test data in return-to-play decisions is contingent upon the integrity of the baseline data for comparison; however, baseline testing may be affected by such variables as learning and attention disorders (Collins et al., 1999; Solomon & Haase, 2008), previous concussions (Collins et al., 1999), and group versus individual test administration (Moser, Schatz, Neidzwski, & Ott, 2011). Effort and perhaps malingering have been shown to affect baseline scores as well. That is, in high school athletes, Hunt, Ferrara, Miller, and Macciocchi (2007) found that 11% of their sample showed poor effort (motivation unknown) on baseline testing, whereas Schatz, Neidzwski, Moser, and Karpf (2010) found that 9% of their sample provided invalid test scores at baseline. In a collegiate sample, Schatz (2010) excluded 6% of his sample for erratic performance. Although Solomon and Haase (2008) found that only 6% of their National Football League (NFL) sample provided invalid test scores, in a recent interview with Peyton Manning of the NFL (Reilly, 2011), Manning reported his intentional poor performance on the NFL's baseline concussion tests in order to not be benched after being injured. These different levels of motivation on baseline testing have been shown to subsequently affect post-concussion neurocognitive test performance (Bailey, Echemendia, & Arnett, 2006), which may negatively influence the quality of return-to-play decisions.

The Immediate Post-concussion Assessment and Cognitive Test (ImPACT Applications, Inc., 2007) is one of several commercially available computerized baseline and post-concussion testing batteries. The reliability and validity of the ImPACT battery has been well established in the assessment of cognitive functions and symptoms associated with concussion (Elbin, Schatz, & Covassin, 2011; Iverson, Gaetz, Lovell, & Collins, 2004; Iverson, Lovell, & Collins, 2005; Lovell et al., 2003; Schatz et al., 2010). It is currently used on professional, collegiate, and high-school athletes. While much attention has

doi:10.1093/arclin/acs050 Advance Access publication on 8 June 2012

been paid by the test creators to assuring the integrity of the baseline test by establishing "red flags" and validity indicators for poor performance, these indicators are generally based on identifying performances 2 *SD* below the mean. It is possible that athletes with motivation to perform poorly on their baseline tests, ostensibly to affect return-to-play decisions later, may significantly lower their baseline scores from where they should be, without reaching threshold on any of the "red flags" or validity indicators.

The purpose of this study was to assess how likely it was that athletes could sandbag their baseline scores without being detected by the "red flags" and validity indicators. In addition, qualitative inquiry was conducted about strategies used to sandbag neurocognitive test performance.

Hypotheses

With no similar data to inform hypotheses, the possibility of significantly lowering baseline scores without being detected was explored by identifying variables which might predict success at "sandbagging" the baseline, such as gender, sport, or concussion history. Participant selection, instruction, and motivation were conservatively controlled in order to limit extraneous variables. Given those parameters, the null hypotheses were tested; that is, that baseline test scores would not be significantly different between motivated and malingered performance and would not be different based on gender, sport, or concussion history.

Method

Participants

Seventy-five undergraduate athletes (33 men and 42 women), who had completed their athletic careers, participated in the study. Sports included men's basketball ($n = 5$), football ($n = 10$), men's lacrosse ($n = 5$), women's lacrosse ($n = 10$), men's soccer ($n = 6$), women's soccer ($n = 9$), softball ($n = 2$), men's swimming/diving ($n = 3$), women's swimming/diving ($n = 1$), men's tennis ($n = 4$), women's tennis ($n = 7$), women's water polo ($n = 6$), and volleyball ($n = 7$). The mean age at baseline exam was 19.99 years ($SD = 1.19$). The mean time lapse from baseline to experimental administration was 17.44 months ($SD = 8.81$).

Materials

ImPACT 6.0 (ImPACT Applications, Inc., 2007) baseline tests were used at both the baseline and experimental administrations. ImPACT 6.0 has six modules (Word Memory, Design Memory, Xs and Os, Symbol Match, Color Match, and Three Letters) which create five composite scores (Verbal Memory, Visual Memory, Processing Speed, Reaction Time, and Impulse Control).

"Red flags" and validity indicators have been identified to assist the test administrator in identifying invalid tests. The four "red flags" include (a) Processing Speed Composite < 25, (b) Reaction Time Composite > 0.8 s, (c) Verbal Memory Composite $< 70\%$, and (d) Visual Memory Composite $< 60\%$. The five validity indicators included (a) Xs and Os Total Interference Incorrect > 30, (b) Impulse Control Composite > 30, (c) Word Memory Learning Percent Correct $< 69\%$, (d) Design Memory Learning Percent Correct $< 50\%$, and (e) Three Letters Total Letters Correct < 8 (ImPACT Applications, Inc., 2007).

Procedure

The Institutional Review Board at the author's institution approved this study. In order to utilize participants who would no longer require taking the ImPACT, participants were recruited via email following either the completion of the seniors' seasons or knowledge of a career-ending injury. Athletes were eliminated from consideration if their athletic careers would likely go beyond college.

The baseline administrations had been conducted in groups by the Athletic Department and took place in campus computer labs with athletes sitting at every other computer. The experimental administrations took place individually in an assessment lab in the Psychology Department.

When participants entered the experimental administration, it was explained to them that the point of the study was to try to perform more poorly on the ImPACT than they had at baseline without reaching threshold on the validity indicators (see Appendix for participant instructions). The participants then gave their informed consent for the study and to use their baseline

data for comparison. The participants were each paid a $5 gift card for participation and, as added incentive in each academic year the study was on-going, the participant who scored the highest on the Impulse Control Composite without reaching threshold was given a $20 gift card.

Results

Quantitative Data

The main independent variable was the context in which the ImPACT was taken, which differed in both when it was taken and under what instruction it was taken, yielding comparisons between the two levels (baseline and experimental administrations). The dependent measures were the nine "red flag" and validity indicator variables. One participant's original baseline data were excluded due to having an Impulse Control Composite > 30, an exclusion criterion typical in real-world baseline test administration (Schatz et al., 2010; Solomon & Haase, 2008). Fifteen participants' baseline administrations were inadvertently but unsystematically deleted by the Athletic Department prior to data analysis, leaving 59 complete and valid baseline administrations and 75 complete experimental administrations. The baseline data of the current sample are consistent with published norms for university men and women (ImPACT, 2012). That is, the current sample's baseline scores on the four published composite variables were all in the Average or High Average range of university students.

Descriptive and inferential statistics of the "red flags" and validity indicators are presented in Table 1. Paired-samples *t*-tests revealed that baseline scores were significantly better than experimental scores on all variables and suggest that participants can in fact score lower on baseline tests when instructed to do so.

Successful faking was defined conservatively as not reaching threshold on any of the "red flags" or validity indicators. Eight participants (11%) were identified as successful fakers, while the remaining 67 were identified as faking (unsuccessful fakers) by one or more of the "red flags" or validity indicators. Men were not overrepresented among the successful fakers, $\chi^2(1, N = 75) = 0.13$, $p = .72$. Participants who played contact sports were not overrepresented among the successful fakers, $\chi^2(1, N = 75) = 2.15$, $p = .14$. Twenty-eight participants self-reported at least one concussion in their histories. There were no significant relationships between number of concussions (range: 0–5) and any of the "red flags" or validity indicators, $ps > .05$. Successful fakers were not overrepresented by those with a concussion history, $\chi^2(1, N = 75) = 0.61$, $p = .43$. The baseline scores of the successful fakers were not significantly different from the baseline scores of the unsuccessful fakers on eight of the nine dependent measures, $ps > .05$. On only Word Memory Learning % Correct did the successful fakers start out with a significantly higher mean ($M = 1.00$, $SD = 0$) than the unsuccessful fakers ($M = 0.98$, $SD = 0.04$), $t(52) = -4.05$, $p < .001$; however, it should be noted that the successful fakers exhibited a ceiling effect with no variability.

Table 2 shows the descriptive and inferential statistics of the "red flags" and validity indicators of the successful fakers. Paired *t*-tests revealed that, for the successful fakers, baseline scores were significantly better than experimental scores on all variables but the Reaction Time Composite and Three Letters Total Letters Correct. Table 3 shows the frequencies and percentages of the participants identified as fakers by each "red flag" and validity indicator. The Verbal Memory Composite and Visual Memory Composite identified the greatest number of participants, and the Reaction Time Composite and Three Letters Total Letters Correct identified the fewest participants.

Table 1. Descriptive and inferential statistics of "red flag" and validity indicator variables in baseline versus experimental administrations

	Baseline ($n = 59$)	Experimental ($n = 75$)	Baseline versus experimental paired-samples *t*-tests and effect sizes
"Red Flags"			
Processing Speed Composite	44.20 (7.10)	26.75 (9.12)	$t(58) = 11.59$, $p < .001$, $r^2 = .71$
Reaction Time Composite	0.52 (0.05)	0.72 (0.22)	$t(58) = -6.32$, $p < .001$, $r^2 = .41$
Verbal Memory Composite	0.92 (0.07)	0.62 (0.14)	$t(58) = 15.90$, $p < .001$, $r^2 = .81$
Visual Memory Composite	0.81 (0.12)	0.51 (0.16)	$t(58) = 12.30$, $p < .001$, $r^2 = .72$
Validity Indicators			
Xs and Os Total Incorrect	8.02 (6.44)	25.83 (23.07)	$t(58) = -5.70$, $p < .001$, $r^2 = .36$
Impulse Control Composite	8.27 (6.45)	31.71 (26.09)	$t(58) = -6.69$, $p < .001$, $r^2 = .44$
Word Memory Learning % Correct	0.98 (0.04)	0.60 (0.18)	$t(58) = 15.96$, $p < .001$, $r^2 = .81$
Design Memory Learning % Correct	0.87 (0.12)	0.57 (0.16)	$t(58) = 14.43$, $p < .001$, $r^2 = .77$
Three Letters Total Letters Correct	14.42 (1.10)	11.43 (2.54)	$t(58) = 8.02$, $p < .001$, $r^2 = .53$

Notes: Sixteen experimental participants did not have baseline data. The paired-sample *t*-tests were performed on $n = 59$.

Table 2. Descriptive and inferential statistics of "red flag" and validity indicator variables in baseline versus experimental administrations of the successful fakers

	Baseline ($n = 6$)	Experimental ($n = 8$)	Baseline versus experimental paired-samples t-tests and effect sizes
"Red flags"			
Processing Speed Composite	42.52 (5.07)	36.88 (5.17)	$t(5) = 6.24$, $p = .002$, $r^2 = .88$
Reaction Time Composite	0.49 (0.04)	0.54 (0.07)	$t(5) = -2.46$, $p = .057$, $r^2 = .55$
Verbal Memory Composite	0.95 (0.05)	0.83 (0.07)	$t(5) = 6.04$, $p = .002$, $r^2 = .88$
Visual Memory Composite	0.85 (0.12)	0.71 (0.06)	$t(5) = 4.43$, $p = .007$, $r^2 = .79$
Validity Indicators			
Xs and Os Total Incorrect	5.00 (3.23)	13.00 (4.87)	$t(5) = -7.81$, $p = .001$, $r^2 = .92$
Impulse Control Composite	5.17 (3.13)	15.25 (4.92)	$t(5) = -8.86$, $p < .001$, $r^2 = .94$
Word Memory Learning % Correct	1.00 (0.0)	0.80 (0.09)	$t(5) = 4.39$, $p = .007$, $r^2 = .79$
Design Memory Learning % Correct	0.94 (0.06)	0.76 (0.08)	$t(5) = 8.76$, $p < .001$, $r^2 = .94$
Three Letters Total Letters Correct	14.00 (1.55)	13.50 (1.20)	$t(5) = 0.0$, $p = 1.000$

Notes: Two successful fakers did not have baseline data. The paired-samples t-tests were performed on $n = 6$.

Table 3. Frequencies and percentages of participants identified as fakers by each "red flag" and validity indicator ($n = 75$)

	n	Percentage
"Red flags"		
Processing Speed Composite	29	39
Reaction Time Composite	15	20
Verbal Memory Composite	55	73
Visual Memory Composite	52	69
Validity Indicator		
Xs and Os Total Incorrect	22	29
Impulse Control Composite	28	37
Word Memory Learning % Correct	48	64
Design Memory Learning % Correct	21	28
Three Letters Total Letters Correct	7	9

In order to determine the probability that the eight participants were identified as successful fakers by chance, the cut-off scores for each "red flag" and validity indicator were calculated as z-scores within the experimental distribution to determine the proportion of the distribution above the cut-off [$z =$ (cut-off score − experimental mean)/experimental standard deviation]. The nine proportions were then multiplied for the joint probability of being above the cut-off on all nine measures, yielding an estimate of 0.15% to have been designated successful fakers by chance. In the current sample, 11% were designated successful fakers, suggesting that their approach was due to something other than chance.

Qualitative Data

After each completed module, the test administrator asked the participants if there were "any techniques or strategies or thought processes that you used to lowball." The responses from the modules whose variables also served as non-composite validity indicators (Word Memory, Design Memory, Xs and Os, and Three Letters) were analyzed by identifying the most frequently reported strategies. Two research assistants independently identified the three most frequent strategies for each variable on the data available to them in consecutive years ($n = 29$ and $n = 57$), with 75% consistency between them. The author independently identified the single highest frequency strategy for each variable on the full sample ($n = 75$), with 100% consistency with the two research assistants' highest frequency strategies. Reporting below is limited to one strategy per variable as the frequencies of the second and third most frequent strategies diminished quickly with 67 unsuccessful participants and quicker still with only eight successful participants.

Word memory learning percent correct. The most frequent strategies of the successful participants (50%) were mild numerical strategies (e.g., getting every 4th answer wrong), while the most frequent strategies of the unsuccessful participants (27%) were numerical strategies yielding scores lower than the cut-off score for detection (e.g., getting all wrong, getting 50% right).

Design memory learning percent correct. There was no discernible pattern in the responses of the successful participants for this module. The most frequent strategies of the unsuccessful participants (21%) were numerical strategies that yielded scores lower than the cut-off score for detection (e.g., alternating yes/no, getting 1 in 5 correct).

Xs and Os total interference incorrect. For the interference task, 25% of the successful participants reported intentionally inaccurate red/blue clicking after a long string of the same color, as did 31% of the unsuccessful participants. The most frequent strategy of the successful participants (25%) was clicking faster in order to naturally facilitate errors in accuracy, but not to directly create inaccurate responses, while the most frequent strategy of the unsuccessful participants (13%) was intentional inaccuracies in the red/blue task.

Three letters total letters correct. The most frequent strategy of the successful participants (38%) was to either do well on the letter memory or the counting, but not both. The most frequent numerical strategies of the unsuccessful participants (15%) were strategies that yielded scores lower than the cut-off score (e.g., remember only one letter). Qualitatively, the unsuccessful participants utilized several types of strategies to perform poorly; mixing up letters with similar letters (30%), and putting the correct letters in a different order (28%).

Discussion

This study revealed that it is difficult for athletes to intentionally perform poorly on the ImPACT without reaching threshold on the "red flags" or validity indicators, as only 11% were able to do it successfully. Those who were successful did not possess any predictive characteristics such as concussion history or sport played; however, qualitative patterns of testing behavior, when compared with the unsuccessful fakers, revealed strategies more focused on natural errors rather than on calculated errors and subtler rather than more flagrant errors.

The Reaction Time Composite and Three Letters Total Letters Correct not only identified the fewest participants as fakers from the full sample, but also were the variables on which it appears the successful fakers chose not to perform poorly. Reaction time has recently been identified as a good predictor of neurocognitive malingering (Bender & Rogers, 2004; Reicker, 2008; Willison & Tombaugh, 2006), as simulators tend to overestimate the effects of head injury. Willison and Tombaugh (2006) reported that their simulator group was unable to simulate realistic reaction times either due to misconceptions about reaction time post-head injury or due to the added cognitive processing time required to malinger their tests. The authors reasoned that once their participants had thought about how many items to get incorrect, they were slower than those who more naturally responded, which appears to be supported by the qualitative reports of the current participants, where the successful fakers reported more strategies related to making natural errors.

It remains somewhat unclear why Three Letters Total Letters Correct identified the fewest participants as fakers and was perhaps not chosen by the successful fakers as a task on which to perform poorly. One aspect of Three Letters Total Letters Correct which is shared by the reaction time tasks is a lack of feedback to the participant, which may contribute to a lack of confidence on how to perform poorly. It is also possible that Three Letters was a test on which it was perceived to be too easy to perform poorly. That is, in contrast to the other modules where the participant is expected to remember up to 12 pieces of information, remembering three letters may seem too simple and transparent a task.

Concussion history did not improve participants' ability to perform more poorly on the ImPACT without detection. This finding is similar to those of Vickery and colleagues (2004) and Dearth and colleagues (2005) who examined the ability of those with moderate-severe head injury to feign head injury symptoms. Vickery and colleagues and Dearth and colleagues found that those with head injury were no more able to feign believable symptoms than non-head injured individuals. The current data indicate that experience with mild head injury (i.e., concussion), like Vickery and colleagues and Dearth and colleagues found with moderate–severe head injury, does not help inform the production of believable cognitive decrements.

The current study did not assess whether coaching strategies would affect ImPACT scores; however, there is an extensive literature to suggest that they may. Rather than coaching symptoms of head injury or cognitive impairment, test-coached simulators have been shown to provide more believable malingered test scores (Powell, Gfeller, Hendricks, & Sharland, 2004; Rogers, 2008). Future research should investigate how multiple baseline exposures (from multiple teams or multiple years) may affect the ability of those with motivation to "sandbag" their current team's baseline. Questioning multiple baselines is in contrast to conventional wisdom that multiple baselines are more reliable than a single baseline (Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007); this idea, however, is predicated upon participants providing their best effort and plateauing practice effects.

Limitations

While the current study showed that it is possible for athletes to "sandbag" their baseline neuropsychological testing, it should be remembered that these participants had exposure to the ImPACT baseline test before and were given basic instruction on the presence of the test's validity indicators. Neither of these conditions would be true for naïve test-takers. Also, while both real-world "sandbaggers" and this study's "sandbaggers" would have external incentive, $25 and easier return-to-play decisions are decidedly different incentives and might be expected to reveal themselves differently on malingered testing.

Lastly, the original baseline tests were administered in groups, while the experimental test was administered individually. Moser and colleagues (2011) found that group administration yielded significantly lower scores than individual administration on several ImPACT measures. There are several factors that may mitigate the impact of this difference in the current study. The current study's baseline tests were administered by athletic trainers with a common script who adhered to all recommendations of administration in the ImPACT manual (ImPACT Applications, Inc., 2007), in contrast to the retrospective group data utilized by Moser and colleagues. While it may be true that individual test administration yields higher scores, this has been found in groups who are ostensibly giving their best effort. The current study's experimental manipulation to explicitly encourage poor performance in the experimental test would effectively negate the premise on which Moser and colleagues found their differences. However, if it is true that group-administered baseline tests yield scores significantly below what the participants are capable (effectively, unintentionally sandbagging), then it is possible that the current study, by defining successful faking only as scores below baseline but above cut-off scores, has underestimated the ability of athletes to sandbag, due to a more limited range in which to perform poorly than those who may have been baseline tested individually.

Conclusion

The current study showed that, while possible, it is difficult for athletes to perform significantly more poorly on their ImPACT baseline than they are capable. Successful fakers tended to use strategies that would produce more natural profiles and they did not perform poorly on reaction time tests and the Three Letters test, suggesting that these tests were seen as either irrelevant to poor performance, too difficult to malinger, or were seen as too obvious a skill to malinger. While personal experience with concussion did not predict successful faking, future research should investigate the impact that coaching poor performance may have on the integrity of neuropsychological baseline testing.

Funding

This work was supported by the Colorado College Psychology Department Sabine Funds.

Acknowledgements

The author would like to thank Sean Guillory, Scarlett Prati, Jordan Evans, and Marina Leith for help in data collection, and the Colorado College Athletic Department for use of their ImPACT program. This paper was accepted for presentation at the 2012 Mid-year Meeting of the International Neuropsychological Society, Oslo Norway, June 27–30, 2012.

Appendix

Participant Instructions

Welcome to the study on how athletes can successfully sandbag a baseline concussion test. We thank you for your participation. You were chosen because of your status as a senior who will shortly be graduating from this institution and also because you were an athlete of the collegiate level. Even though you are practically on your way out as a student athlete, we ask you to take one more test. The point of this test is to fail though. To make it even more strange, we ask you to try to fail a test that you have already taken.

The test in question is the ImPACT Concussion Battery to test your baseline cognitive abilities in the unfortunate event that you may have gotten a concussion. We hope that you were in the group that tried their hardest the last time you took it so a valid assessment could be made on your condition; unfortunately, this is not always the case. It is quite possible to sandbag this baseline test so any successive tests to follow aren't properly measuring cognitive deficits induced from head injury. This could mean that someone could have a serious brain injury and the test results could say that it is OK for them to play. The test does have some ways of checking if someone is sandbagging, but it does not always catch it.

For this study, we want you to try to beat the system by trying to get as poor scores as possible while not getting identified by the test that you are indeed not giving complete effort. How do you do that? We would like you to try any technique, strategy, or mental trick that may help you beat the test. Whatever you do, we will ask that you try to describe your thoughts and behaviors that you used to try to do so. We will ask you about your effort and these techniques used after every module.

The incentive is that if you are the top sandbagger without being detected, your prize will be a $20 gift certificate in addition to the $5 gift card for just participating. Now, it is time to begin the test. Good luck. Any questions?

References

Bailey, C., Echemendia, R., & Arnett, P. (2006). The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. *Journal of the International Neuropsychological Society*, *12*, 475–484.

Bender, S., & Rogers, R. (2004). Detection of neurocognitive feigning: Development of a multi-strategy assessment. *Archives of Clinical Neuropsychology*, *19*, 49–60.

Broglio, S., Ferrara, M., Macciocchi, S., Baumgartner, T., & Elliott, R. (2007). Test-retest reliability of computerized concussion assessment programs. *Journal of Athletic Training*, *42 (4)*, 509–514.

Collins, M., Grindel, S., Lovell, M., Dede, D., Moser, D., Phalin, B., et al. (1999). Relationship between concussion and neuropsychological performance in college football players. *Journal of the American Medical Association*, *282 (10)*, 964–970.

Dearth, C., Berry, D., Vickery, C., Vagnini, V., Baser, R., Orey, S., et al. (2005). Detection of feigned head injury symptoms on the MMPI-2 in head injured patients and community controls. *Archives of Clinical Neuropsychology*, *20*, 95–110.

Elbin, R., Schatz, P., & Covassin, T. (2011). One-year test-retest reliability of the online version of ImPACT in high school athletes. *American Journal of Sports Medicine*, *39 (11)*, 2319–2324.

Hunt, T., Ferrara, M., Miller, L., & Macciocchi, S. (2007). The effect of effort on baseline neuropsychological test scores in high school football athletes. *Archives of Clinical Neuropsychology*, *22*, 615–621.

ImPACT Applications, Inc. (2007). *ImPACT 2007 (6.0) Clinical Interpretation Manual*. Pittsburgh, PA.

ImPACT. (2012). Normative Baseline Data. Retrieved April 18, 2012, from http://impacttest.com/publications/baseline_data/normative.

Iverson, G., Gaetz, M., Lovell, M., & Collins, M. (2004). Relation between subjective fogginess and neuropsychological testing following concussion. *Journal of the International Neuropsychological Society*, *10*, 904–906.

Iverson, G., Lovell, M., & Collins, M. (2005). Validity of ImPACT for measuring processing speed following sports-related concussion. *Journal of Clinical and Experimental Neuropsychology*, *27*, 683–689.

Kaminski, T., Groff, R., & Glutting, J. (2009). Examining the stability of Automated Neuropsychological Assessment Metric (ANAM) baseline test scores. *Journal of Clinical and Experimental Neuropsychology*, *31 (6)*, 689–697.

Lovell, M., Collins, M., Iverson, G., Field, M., Maroon, J., Cantu, R., et al. (2003). Recovery from mild concussion in high school athletes. *Journal of Neurosurgery*, *98*, 295–301.

Moser, R. S., Schatz, P., Neidzwski, K., & Ott, S. (2011). Group versus individual administration affects baseline neurocognitive test performance. *American Journal of Sports Medicine*, *39*, 2325–2330.

Powell, M., Gfeller, J., Hendricks, B., & Sharland, M. (2004). Detecting symptom- and test-coached simulators with the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, *19*, 693–702.

Reicker, L. (2008). The ability of reaction time tests to detect simulation: An investigation of contextual effects and criterion scores. *Archives of Clinical Neuropsychology*, *23*, 419–431.

Reilly, R. (2011). *Talking football with Archie, Peyton, Eli*. Retrieved April 29, 2011, from http://sports.espn.go.com/espn/news/story?id=6430211.

Rogers, R. (2008). Researching response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.). New York: The Guilford Press.

Schatz, P. (2010). Long-term test-retest reliability of baseline cognitive assessments using the ImPACT. *American Journal of Sports Medicine*, *38 (1)*, 47–52.

Schatz, P., Neidzwski, K., Moser, R. S., & Karpf, R. (2010). Relationship between subjective test feedback provided by high-school athletes during computer-based assessment of baseline cognitive functioning and self-reported symptoms. *Archives of Clinical Neuropsychology*, *25*, 285–292.

Solomon, G., & Haase, R. (2008). Biopsychosocial characteristics and neurocognitive test performance in National Football League players: An initial assessment. *Archives of Clinical Neuropsychology*, *23*, 563–577.

Vickery, C., Berry, D., Dearth, C., Vagnini, V., Baser, R., Cragar, D., et al. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology*, *19*, 37–48.

Willison, J., & Tombaugh, T. (2006). Detecting simulation of attention deficits using reaction time tests. *Archives of Clinical Neuropsychology*, *21*, 41–52.