

Interpreting Change on ImPACT Following Sport Concussion

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

¹Department of Psychiatry, University of British Columbia & Riverview Hospital, Vancouver, BC, Canada, and ²Department of Orthopaedic Surgery, University of Pittsburgh Medical Center, Sports Medicine Concussion Program, Pittsburgh, PA, USA

ABSTRACT

The purpose of this study was to examine the psychometric characteristics of Version 2.0 of ImPACT (Immediate Postconcussion Assessment and Cognitive Testing). The focus was on the stability of the test scores and the calculation of reliable change confidence intervals for the test-retest difference scores. A sample of 56 nonconcussed adolescents and young adults completed the test battery on two occasions. Test-retest coefficients, reliable change difference scores, and confidence intervals for measurement error are provided. These reliable change parameters were applied to a second sample of 41 concussed amateur athletes who were tested preseason and within 72 hr of injury. Applying these confidence intervals allows more precise determinations of deterioration, improvement, and recovery in the initial days following concussion.

Estimating change is the sine qua non of clinical neuropsychology. Every neuropsychological evaluation includes a careful determination of change. Typically, we try to estimate decline in functioning that can be attributed to a brain injury, condition, or disease. Other evaluations are undertaken to assess interval change. Neuropsychological assessment can be very useful for tracking recovery from a traumatic brain injury or a stroke, or for monitoring progression of a dementing disease such as Alzheimer's. For practical, clinical, and economic reasons, follow-up evaluations typically are conducted after 6–24 months (although, there are clinical situations in which shorter retest intervals are preferred).

Sports neuropsychology is relatively unique in that cognitive assessments often occur over very brief retest intervals to facilitate decisions regarding returning to practice and competition. This creates special challenges relating to estimating

change. For example, the phenomenon under study, the effects of concussion on cognitive functioning, is rapidly changing. The accuracy with which we can assess this phenomenon is related to the sensitivity of the measures, and, of course, their reliability.

ImPACT (Immediate Postconcussion Assessment and Cognitive Testing) is a computerized neuropsychological screening battery designed specifically for assessing sports-related concussion. Version 1.0 of the battery has been used in several studies relating to outcome from concussion (Collins et al., 2003; Iverson, Gaetz, Lovell, & Collins, 2002a; Iverson, Gaetz, Lovell, Collins, & Maroon, 2002b; Lovell et al., 2003; Lovell, Collins, Iverson, Johnston, & Bradley, in press). The battery was designed to minimize practice effects through the use of several alternate forms. In a reliability study for Version 1.0, there were no practice effects over a 2-week retest interval in a

sample of 49 amateur athletes (Iverson, Lovell, Collins, & Norwig, 2002c). Reliable change estimates for Version 1.0 were provided.

The purpose of this study is to provide detailed information regarding the interpretation of change on Version 2.0 of ImPACT. Test scores can be influenced by numerous factors, such as practice effects, regression to the mean, and more random or unpredictable forms of measurement error. Therefore, proper interpretation of the test requires an understanding of the probable range of measurement error that surrounds test-retest difference scores. This allows more precise determinations of deterioration, improvement, and recovery in the initial days following concussion. First, test-retest reliability, practice effects, and reliable change parameters will be estimated in a sample of healthy young people who completed the battery over a brief retest interval (i.e., approximately 7 days). Second, the derived reliable change parameters will be applied to a sample of amateur athletes who underwent pre-season testing and were re-evaluated within 72 hr of sustaining a concussion.

METHOD

Participants and Procedures

The first sample was comprised of 56 adolescents and young adults who completed Version 2.0 of ImPACT twice for the purpose of a test-retest study. There were 29 males and 27 females. Their average age was 17.6 years ($SD = 1.7$, range = 15–22). Approximately 64% were in high school and 36% were in university. The average retest interval was 5.8 days (median = 7, $SD = 3.0$, range = 1–13). Approximately 29% were retested within 3 days, 43% within 4 days, 82% within 7 days, and 95% within 11 days.

The second sample was comprised of 41 amateur athletes who sustained a sports-related concussion. All athletes completed ImPACT at the beginning of the season. All were retested within 72 hr of their concussions (mean = 1.3, median = 1, $SD = 0.7$ days). This sample was 90% male. Their average age was 16.8 years (median = 16, $SD = 2.4$, range = 13–22). Approximately 71% were in high school and 29% were in university. The vast majority of athletes were football players (88%), with small numbers of athletes in other sports such as hockey, soccer, basketball, and wrestling. Most athletes had sufficient information to classify the severity of their concussions using the

American Academy of Neurology Concussion Grading System (Kelly & Rosenberg, 1998; Quality Standards Subcommittee, 1997). Approximately 54% had Grade I Concussions, 22% had Grade II Concussions, and 7% had Grade III Concussions. Missing data prevented the confident classification of 17% (i.e., 7 athletes).

Measure

Version 2.0 of ImPACT is a computer administered neuropsychological test battery that consists of six individual test modules that measure aspects of cognitive functioning including attention, memory, reaction time, and processing speed. Four composite scores were used for this study. In general, the test battery is designed to yield multiple types of information within a brief period of time. Each test module may contribute scores to multiple composite scores. The Verbal Memory composite score represents the average percent correct for a word recognition paradigm, a symbol number match task, and a letter memory task with an accompanying interference task. The Visual Memory composite score is comprised of the average percent correct scores for two tasks; a recognition memory task that requires the discrimination of a series of abstract line drawings, and a memory task that requires the identification of a series of illuminated X's or O's after an intervening task (mouse clicking a number sequence from 25 to 1). The Reaction Time composite score represents the average response time (in milliseconds) on a choice reaction time, a go/no-go task, and the previously mentioned symbol match task. The Processing Speed composite represents the weighted average of three tasks that are done as interference tasks for the memory paradigms. The Impulse Control composite score represents the total number of errors of omission or commission on the go/no-go test and the choice reaction time test. This composite is used to identify athletes who are not putting forth maximum effort or who are seriously confused about test instructions. This composite was not one of the dependent measures for this study. In addition to the cognitive measures, ImPACT also contains a Postconcussion Symptom Scale, utilized throughout organized sports (Aubry et al., 2002; Lovell & Collins, 1998), that consists of 21 commonly reported symptoms (e.g., headache, dizziness, "fogginess"). The dependent measure is the total score derived from this 21-item scale.

Most research to date has used version 1.0 of the program. ImPACT 2.0 is very similar to the original version. However, there are some significant changes. Version 2.0 includes an additional test module (design memory). In addition, one of the working memory tasks (X's and O's) was expanded and modified, making it more difficult than the previous version. Version 2.0

also yields two memory composite scores (Verbal Memory and Visual Memory) while Version 1.0 contains only one memory composite score.

Design and Analysis

The first set of analyses were based on the healthy young people tested twice. This was a within subjects design. Relative position across the two distributions was examined with a Pearson correlation. Level of performance within subjects was examined with dependent *t*-tests. Reliable change estimates were derived from a modification of the method proposed by Jacobson and Truax (1991). This methodology has been used extensively in clinical psychology (Hageman & Arrindell, 1993; Hsu, 1989; Jacobson & Revenstorf, 1988; Jacobson, Roberts, Berns, & McGlinchey, 1999; Ogles, Lambert, & Masters, 1996; Speer, 1992; Speer & Greenbaum, 1995), clinical neuropsychology (Chelune, Naugle, Luders, Sedlak, & Awad, 1993; Heaton et al., 2001; Iverson, 1998, 1999, 2001; Temkin, Heaton, Grant, & Dikmen, 1999), and sports neuropsychology (Barr & McRea, 2001; Hinton-Bayre, Geffen, Geffen, McFarland, & Friis, 1999; Iverson et al., 2002c). The reliable change methodology allows the clinician to estimate measurement error surrounding test-retest difference scores. Specifically, the standard error of difference (S_{diff}) is used to create a confidence interval for the baseline-retest difference score. The steps for calculating the S_{diff} are provided below:

$$SEM_1 = SD\sqrt{1 - r_{12}}$$

(Standard deviation from time 1 multiplied by the square root of 1 minus the test-retest coefficient).

$$SEM_2 = SD\sqrt{1 - r_{12}}$$

(Standard deviation from time 2 multiplied by the square root of 1 minus the test-retest coefficient).

$$S_{diff} = \sqrt{SEM_1^2 + SEM_2^2}$$

(Square root of the sum of the squared SEMs for each testing occasion).

The reader should note that the formula used in this study for calculating the S_{diff} uses the *SEM* for baseline and retest, whereas many past studies have used an “estimated” S_{diff} by simply multiplying the squared baseline *SEM* by two (i.e., $\sqrt{2SEM_1^2}$). The estimated S_{diff} should only be used when retest data are not available (Hageman & Arrindell, 1993; Iverson, 1998, 2001). Several refinements and modifications to the reliable change methodology have been debated in the literature (Hageman & Arrindell, 1993, 1999a, 1999b; Hsu, 1989, 1999; Speer, 1992; Speer & Greenbaum, 1995). The issues are far from resolved. We chose to use the reliable change method that corrects for practice (Chelune et al., 1993; Iverson & Green, 2001), when practice effects are present.

RESULTS

Descriptive statistics for the healthy young people tested twice are presented in Table 1. The Pearson test-retest correlation coefficients for the composite scores were as follows: Verbal Memory = 0.70, Visual Memory = 0.67, Reaction Time = 0.79, Processing Speed = 0.86, and Post-concussion Scale = 0.65. The standard errors of measurement (*SEMs*), standard errors of difference (S_{diff} s), and reliable change confidence intervals also are presented in Table 1. The probable ranges of measurement error for the ImpACT composites are as follows: Verbal Memory Composite = 6.83 points, Visual Memory Composite = 10.59 points, Reaction Time Composite = 0.05 s, Processing Speed

Table 1. Descriptive Statistics, *SEMs*, S_{diff} s, and Reliable Change Confidence Intervals for the Healthy Control Subjects (*N* = 56).

Composite	<i>M</i> (<i>SD</i>)		<i>p</i>	<i>SEM</i> ₁	<i>SEM</i> ₂	S_{diff}	Confidence intervals	
	Time 1	Time 2					0.80	0.90
Verbal Memory	88.68 (9.50)	88.84 (8.09)	.86	5.20	4.43	6.83	8.75	11.21
Visual Memory	78.70 (13.39)	77.48 (12.67)	.40	7.69	7.28	10.59	13.55	17.37
Reaction Time	.543 (.087)	0.536 (.063)	.34	0.04	0.03	0.05	0.06	0.08
Processing Speed	40.54 (7.64)	42.24 (7.06)	.002	2.86	2.64	3.89	4.98	6.38
Postconcussion Scale	5.23 (6.75)	5.79 (10.07)	.59	3.99	5.96	7.17	9.18	11.76

Note. *SEM*: standard error of measurement; S_{diff} : standard error of difference.

Composite = 3.89 points, and Postconcussion Scale 7.17 points. The 80% confidence intervals for estimating change are as follows: Verbal Memory ≥ 9 points, Visual Memory ≥ 14 points, Reaction Time > 0.06 s, Processing Speed ≥ 5 points, and Postconcussion Total Scores ≥ 10 points. These are rounded values derived from Table 1.

Level of performance was compared using paired samples *t*-tests. There were no within group differences for Verbal Memory, $t(55) = -0.17, p < .87$, Visual Memory, $t(55) = 0.85, p < .40$, Reaction Time, $t(55) = 0.97, p < .34$, or total symptoms, $t(55) = -0.54, p < .60$. There was a significant difference between baseline and retest on the Processing Speed Composite, $t(55) = -3.26, p < .003, d = 0.23$, small effect size. On average, there was a 1.7 point practice effect for the Processing Speed Composite. Approximately 68% of the sample was faster at retest than at baseline.

The reliable change difference scores associated with the two confidence intervals were applied to the original data. If the distributions of difference scores were perfectly normal, then one would expect to see 10% in each tail for the 0.80 confidence interval and 5% in each tail for the 0.90 confidence interval. As seen in Table 2, the percentages of subjects that would be classified as reliably improved or declined was reasonably close to what would be predicted from the theoretical normal distribution.

The number of scores that reliably declined for each subject was computed. A decline was defined as reliably lower Verbal or Visual memory, slower processing speed or reaction time, or

greater symptoms at retest versus baseline (80% confidence interval). The percentages of subjects showing declines across the five composite scores are as follows: no declines = 63.0%, one decline = 39.3%, two declines = 1.8%, 3 declines = 0%, and 4 declines = 1.8%.

The sensitivity of the composite scores to the acute effects of concussion was estimated in the sample of 41 amateur athletes who were tested preseason and within 72 hr of injury. The athletes demonstrated a significant decline in Verbal Memory (baseline $M = 84.9, SD = 7.2$; Postconcussion $M = 76.8, SD = 12.6; p < .0002, d = 0.82$, large effect size) and Visual Memory (baseline $M = 75.7, SD = 12.3$; Postconcussion $M = 66.4, SD = 14.7; p < .0002, d = 0.69$, medium-large effect size). They also demonstrated significantly slower Processing Speed (baseline $M = 36.9, SD = 6.8$; Postconcussion $M = 33.1, SD = 8.8; p < .006, d = 0.49$, medium effect size), and Reaction Time (baseline $M = 0.56, SD = 0.08$; Postconcussion $M = 0.65, SD = 0.11; p < .00005, d = 0.95$, large effect size). The athletes also demonstrated a large increase in symptom reporting (baseline $M = 8.2, SD = 10.7$; Postconcussion $M = 24.3, SD = 21.7; p < .00001, d = 0.99$, large effect size). These findings are illustrated in Figure 1.

The 80% confidence interval for estimating reliable change was applied to each of the concussed athlete's composite scores. The confidence interval for Processing Speed was adjusted by two points for the presumed practice effect. The breakdown of reliable change for each composite score was as follows: Verbal Memory 44% declined, 7.3% improved; Visual Memory 41.5% declined, 2.4% improved; Reaction Time 51.2%

Table 2. Percentages of the Healthy Sample that Would be Classified as Reliably Improved or Declined Based on the 0.80 and 0.90 Confidence Intervals.

	0.80 confidence interval		0.90 confidence interval	
	Declined (%)	Improved (%)	Declined (%)	Improved (%)
Verbal Memory	10.7	16.1	5.4	8.9
Visual Memory	10.7	8.9	5.4	3.6
Reaction Time	8.9	14.3	5.4	7.1
Processing Speed ^a	7.1	8.9	3.6	5.4
Postconcussion Scale	12.5	7.1	10.7	3.6

Note. ^aThe confidence intervals for the Processing Speed composite were adjusted for a 2-point practice effect.

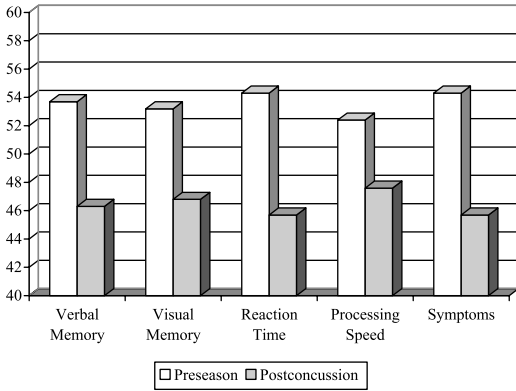


Fig. 1. Comparison of preseason and postinjury scores on the five composites transformed into uniform *T*-scores (*N* = 41).

Note. These *T*-scores are not normative *T*-scores. They are standardized scores. The distributions of baseline and postconcussion scores for each composite were standardized with a mean of 50 and a standard deviation of 10. The direction of the symptom score and the reaction time score was reversed, so that lower *T*-scores represent worse scores. Thus, all five composites can be compared graphically on a common metric.

declined, 7.3% improved; Processing Speed 41.5% declined, 4.9% improved; Postconcussion Scale 53.7% reported more symptoms, 2.4% reported fewer symptoms.

The number of scores that reliably declined for each subject was computed. A decline was defined in the same manner as it was for the healthy test-retest sample. The percentages of athletes showing declines across the five composite score are as follows: no declines = 24.4%, one decline = 12.2%, two declines = 14.6%, three declines = 17.1%, four declines = 19.5%, and five declines = 12.2%. Athletes with concussions are much more likely to have two or more declines across the five composites than the healthy subjects [63.4% vs. 3.6%; $\chi^2(1, 97) = 41.3, p < .00001$; Odds Ratio = 46.8, 95% CI = 10.0–220.0].

DISCUSSION

This study illustrates important aspects of the psychometric properties of Version 2.0 of

ImPACT. The test-retest coefficients for the five composite scores ranged from 0.65 to 0.86. Although, seemingly relatively modest, these stability coefficients are comparable or higher than many other neuropsychological tests, such as the Wechsler Memory Scale – Third Edition Index scores (Psychological Corporation, 1997), Delis–Kaplan Executive Function System Trail-Making Test or Color-Word Test (Delis, Kaplan, & Kramer, 2001), or the California Verbal Learning Test–Second Edition (Delis, Kramer, Kaplan, & Ober, 2000).

When evaluating changes in cognitive performance following concussion, it is critically important to understand the probable range of measurement error surrounding test-retest difference scores to more accurately document deterioration from preseason testing and recovery during the initial days postinjury. In the present study, we made adjustments to the ImPACT Processing Speed composite score reliable change indices because practice effects were present. It was not necessary to adjust the other composite scores because practice effects were not identified. ImPACT was designed to reduce practice effects through randomization of stimuli presentation. This was an essential design feature because the battery is intended to be used repeatedly, over short intervals. A quick reference guide for estimating change on the composite scores is presented in Table 3.

In the second part of this study, preseason and postconcussion scores were examined for 41 concussed amateur athletes. As a group, these athletes demonstrated a large change in Verbal Memory, reaction time, and self-reported symptoms. They experienced a medium-to-large change in Visual Memory and processing speed.

The effect sizes from preseason to postconcussion were medium to large, ranging from 0.49 to

Table 3. Quick Reference Reliable Change Estimates: 80% Confidence Interval.

Composite	Declined	Improved
Verbal Memory	9 points	9 points
Visual Memory	14 points	14 points
Reaction Time	0.06 s	0.06 s
Processing Speed	3 points	7 points
Postconcussion Scale	10 points	10 points

0.99 across the five composite scores. These effect sizes are comparable to the magnitude of "impairments" on other tests in other populations. For example, the effect sizes comparing orthopedically injured trauma control subjects to patients with moderate-severe traumatic brain injuries were 0.46 for the Category Test and 0.37 for Trails B (calculated from Dikmen, Ross, Machamer, & Temkin, 1995; patients with TBIs took 7–13 days postinjury to reliably follow commands, and they were tested 1-year postinjury). Patients with Alzheimer's disease showed a 0.79 effect size for the WAIS-III Working Memory Index and a 1.39 effect size for the Processing Speed Index (Psychological Corporation, 1997).

When the reliable change methodology was applied to the concussed athletes, 44%–54% showed statistically reliable declines across the five individual composite scores. Athletes with concussions were 47 times more likely to have 2 or more declines across the five composites than nonconcussed subjects tested twice. Clearly, the computerized screening battery is sensitive to the acute effects of concussion and a large percentage of athletes show substantial changes in functioning in the first few days postinjury. This sensitivity to the acute effects of concussion is consistent with research with version 1.0 of ImPACT (Collins et al., 2003; Iverson et al., 2002a; Lovell et al., 2003; Lovell et al., in press). It is important to emphasize that concussion is a highly individualized injury. Some athletes experience immediate, pronounced problems whereas others experience very mild problems that resolve quickly. All athletes are not expected to show cognitive problems on neuropsychological testing, even in the first couple days postinjury.

This was a preliminary study designed to investigate reliable change on Version 2.0 of ImPACT. It is limited by the relatively small sample size, a common limitation with most (e.g., Barr, 2003; Hinton-Bayre et al., 1999; Moritz, Iverson, & Woodward, in press; Sawrie, Chelune, Naugle, & Luders, 1996), but not all (e.g., Erlanger et al., 2003; Temkin et al., 1999) reliable change studies. The effect of the heterogeneity of the sample (i.e., high school and college students) on the test-retest coefficients is unknown. Future research with larger, more

homogeneous samples might further refine the interpretation of change on this battery.

Another limitation in this study is the retest interval. This interval was very short. Thus, it is relevant for postconcussion testing over at least one short interval. However, it is possible that the reliable change estimates would change over a longer interval, such as from preseason to postconcussion. This limits the external validity of these results because the brief retest interval in healthy subjects was used to estimate reliable change in healthy then concussed athletes tested at a longer interval. It is also possible that the practice effect seen on the Processing Speed composite might diminish or disappear over a longer retest interval.

Three practical methodological issues relating to estimating reliable change will be presented. First, there is the statistical issue of regression to the mean and the practical issue of an unusually good or unusually poor performance. As a general rule, extreme scores are likely to be less extreme at retest. The reliable change methodology essentially averages this phenomenon into the measurement error estimate. The end result is that the reliable change estimate is optimized for the entire sample but is not as accurate for subsamples, such as the top 20%, middle 60%, and bottom 20% of scores. In other words, one of the most important predictors of a retest score is the level of the baseline score (Sawrie et al., 1996; Temkin et al., 1999). Optimally, reliable change estimates would be based on large samples of more homogeneous baseline scores.

Second, it is most common to present 90% or 95% confidence intervals for reliable change. This is a sensitivity and specificity issue. Do we really want to be 95% sure that the change observed is not due to possible measurement error, leaving only 2.5% in each tail? Under many clinical circumstances we want to adopt a more liberal statistical criterion so that we are more likely to identify real change when it occurs. That is why the 80% confidence interval was emphasized in this study and in previous work (Iverson, 1999, 2001; Iverson & Green, 2001). Barr (2003) recently included the 70% confidence interval.

Third, the issue of practice effects is important (Chelune et al., 1993), yet complicated. Is it appropriate to correct all scores for an "average"

practice effect? What if 55% of subjects score higher at retest and 15% score substantially higher? Correcting for the average practice effect might not be optimal for a large percentage of these subjects. Iverson and Green (2001) recommended correcting for the average practice effect if 75% or more of the sample had a higher score, of any magnitude, at retest. This approach, of course, has limitations too, and more research, especially with regression modeling of large representative samples over relevant retest intervals, is needed.

With regard to the use of neuropsychological assessment procedures in sports medicine, it is important to stress that the reliable change difference scores are meant to supplement, not replace, clinical judgment. The determination of decline and then subsequent improvement in functioning following concussion is a complex clinical process that involves multiple sources of data. This reliable change methodology simply allows clinicians to estimate the probable range of measurement error surrounding test-retest difference scores. Obviously, it is possible for athletes to experience real decline or improvement even if their scores do not exceed the 0.80 confidence interval for measurement error. The practitioner simply should have less confidence in clinical inferences based on changes that fall within the probable range of measurement error, and seek more ancillary evidence to support his or her opinion.

ACKNOWLEDGMENTS

The authors thank Jennifer Bernardo for assistance with manuscript preparation. Additional information regarding IMPACT is available at www.impacttest.com.

REFERENCES

- Aubry, M., Cantu, R., Dvorak, J., Graf-Baumann, T., Johnston, K.M., Kelly, J., Lovell, M.R., McCrory, P., Meeuwisse, W., & Schamasch, P. (2002). Summary and agreement statement of the first international conference on concussion in sport, Vienna, 2001. *Clinical Journal of Sports Medicine, 12*, 6–11.
- Barr, W.B. (2003). Neuropsychological testing of high school athletes: Preliminary norms and test-retest indices. *Archives of Clinical Neuropsychology, 18*, 91–101.
- Barr, W.B., & McCrea, M. (2001). Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. *Journal of the International Neuropsychological Society, 7*, 693–702.
- Chelune, G.J., Naugle, R.I., Luders, H., Sedlak, J., & Awad, I.A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*, 41–52.
- Collins, M.W., Field, M., Lovell, M.R., Iverson, G.L., Johnston, K.M., Maroon, J., & Fu, F.H. (2003). Relationship between post-concussion headache and neuropsychological test performance in high school athletes. *American Journal of Sports Medicine, 31*, 168–173.
- Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis Kaplan Executive Function System technical manual*. San Antonio, TX: The Psychological Corporation.
- Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (2000). *California Verbal Learning Test Second Edition Adult Version manual*. San Antonio, TX: The Psychological Corporation.
- Dikmen, S.S., Ross, B.L., Machamer, J.E., & Temkin, N.R. (1995). One year psychosocial outcome in head injury. *Journal of the International Neuropsychological Society, 1*, 67–77.
- Erlanger, D., Feldman, D., Kutner, K., Kaushik, T., Kroger, H., Festa, J., Barth, J., Freeman, J., & Broshek, D. (2003). Development and validation of a web-based neuropsychological test protocol for sports-related *return-to-play* decision-making. *Archives of Clinical Neuropsychology, 18*, 293–316.
- Hageman, W.J.J.M., & Arrindell, W.A. (1993). A further refinement of the reliable change (RC) index by improving the pre-post difference score: Introducing RC(ID). *Behaviour Research and Therapy, 31*, 693–700.
- Hageman, W.J.J.M., & Arrindell, W.A. (1999a). Clinically significant and practical! Enhancing precision does make a difference. Reply to McGlinchey and Jacobson, Hsu, and Speer. *Behaviour Research and Therapy, 37*, 1219–1233.
- Hageman, W.J.J.M., & Arrindell, W.A. (1999b). Establishing clinically significant change: Increment of precision and the distinction between individual and group level of analysis. *Behaviour Research and Therapy, 37*, 1169–1193.
- Heaton, R.K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M.J., Marcotte, T.D., & Grant, I. (2001). Detecting change: A comparison of three neuropsychological methods using normal and clinical samples. *Archives of Clinical Neuropsychology, 16*, 75–91.

- Hinton-Bayre, A.D., Geffen, G.M., Geffen, L.B., McFarland, K.A., & Friis, P. (1999). Concussion in contact sports: Reliable change indices of impairment and recovery. *Journal of Clinical and Experimental Neuropsychology*, *21*, 70–86.
- Hsu, L.M. (1989). Reliable changes in psychotherapy: Taking into account regression toward the mean. *Behavioral Assessment*, *11*, 459–467.
- Hsu, L.M. (1999). A comparison of three methods of identifying reliable and clinically significant client changes: Commentary on Hageman and Arrindell. *Behaviour Research and Therapy*, *37*, 1195–1202.
- Iverson, G.L. (1998). Interpretation of Mini-Mental State Examination scores in community-dwelling elderly and geriatric neuropsychiatry patients. *International Journal of Geriatric Psychiatry*, *13*, 661–666.
- Iverson, G.L. (1999 July/August). Interpreting change on the WAIS-III/WMS-III in persons with traumatic brain injuries. *Journal of Cognitive Rehabilitation*, *16*–20.
- Iverson, G.L. (2001). Interpreting change on the WAIS-III/WMS-III in clinical samples. *Archives of Clinical Neuropsychology*, *16*, 183–191.
- Iverson, G.L., Gaetz, M., Lovell, M.R., & Collins, M.W. (2002a). Relation between foginess and outcome following concussion. *Archives of Clinical Neuropsychology*, *17*, 769–770.
- Iverson, G.L., Gaetz, M., Lovell, M.R., Collins, M., & Maroon, J. (2002b). Cumulative effects of concussion in amateur athletes. *Archives of Clinical Neuropsychology*, *17*, 769.
- Iverson, G.L., & Green, P. (2001). Measuring improvement or decline on the WAIS-R in inpatient psychiatry. *Psychological Reports*, *89*, 457–462.
- Iverson, G.L., Lovell, M.R., Collins, M.W., & Norwig, J. (2002c). Tracking recovery from concussion using ImPACT: Applying reliable change methodology. *Archives of Clinical Neuropsychology*, *17*, 770.
- Jacobson, N.S., & Revenstorf, D. (1988). Statistics for assessing the clinical significance of psychotherapy issues: Issues, problems, and new developments. *Behavioral Assessment*, *10*, 133–145.
- Jacobson, N.S., Roberts, L.J., Berns, S.B., & McGlinchey, J.B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, *67*, 300–307.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.
- Kelly, J.P., & Rosenberg, J.H. (1998). The development of guidelines for the management of concussion in sports. *Journal of Head Trauma Rehabilitation*, *13*, 53–65.
- Lovell, M.R., & Collins, M.W. (1998). Neuropsychological assessment of the college football player. *Journal of Head Trauma Rehabilitation*, *13*, 9–26.
- Lovell, M.R., Collins, M.W., Iverson, G.L., Field, M., Maroon, J., Cantu, R., Podell, K., Powell, J., & Fu, F.H. (2003). Recovery from mild concussion in high school athletes. *Journal of Neurosurgery*, *98*, 296–301.
- Lovell, M.R., Collins, M.W., Iverson, G.L., Johnston, K.M., & Bradley, J.P. (in press). Grade 1 or “Ding” concussions in high school athletes. *American Journal of Sports Medicine*.
- Moritz, S., Iverson, G.L., & Woodward, T.S. (2003). Reliable change indices for memory performance in schizophrenia as a means to determine drug-induced cognitive decline. *Applied Neuropsychology*, *10*, 115–120.
- Ogles, B.M., Lambert, M.J., & Masters, K.S. (1996). *Assessing outcome in clinical practice*. Boston: Allyn and Bacon.
- Quality Standards Subcommittee. (1997). Practice parameter: The management of concussion in sports (summary statement). *Neurology*, *48*, 581–585.
- Sawrie, S.M., Chelune, G.J., Naugle, R.I., & Luders, H.O. (1996). Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. *Journal of the International Neuropsychological Society*, *2*, 556–564.
- Speer, D.C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology*, *60*, 402–408.
- Speer, D.C., & Greenbaum, P.E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology*, *63*, 1044–1048.
- Temkin, N.R., Heaton, R.K., Grant, I., & Dikmen, S.S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*, *5*, 357–369.
- The Psychological Corporation. (1997). *WAIS-III WMS-III technical manual*. San Antonio, TX: Author.