Background: Computer-based tests are widely used for the purpose of documenting baseline neurocognitive function, and athletes are often tested in groups. At present, there is limited research on the effects of administering these tests in a group versus individual setting.

Hypothesis: Athletes assessed with neurocognitive tests in groups would show decreased performance compared with those assessed individually.

Study Design: Cohort study; Level of evidence, 3.

Methods: High school athletes completed preseason baseline neurocognitive tests either in groups in their school or individually in a private neuropsychological clinic (with no significant between-group differences in sex, history of concussion, and attention deficit disorder/learning disability). All athletes completed the ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing) test battery on 1 occasion, which yielded scores in the area of verbal memory, visual memory, motor processing speed, reaction time, and symptom scores.

Results: Athletes in the group setting scored significantly lower on verbal memory (mean 83.4 vs 86.5; \( P = .003 \)), visual memory (mean 71.6 vs 76.7; \( P = .0001 \)), motor processing speed (mean 35.6 vs 38.4; \( P = .0001 \)), and reaction time (mean 0.61 vs 0.57; \( P = .001 \)), but not symptom scores (mean 6.1 vs 4.4; \( P = .11 \)), and exhibited a greater rate of invalid baselines. Similar results were obtained after excluding athletes with attention deficit disorder/learning disability and/or a history of concussion.

Conclusion: Administering baseline neurocognitive testing to athletes in a group setting may introduce extraneous error, negatively affecting test performance.

Keywords: baseline testing; neurocognitive testing; concussion testing; computerized testing; group administration

Concussion assessment programs commonly utilize computer-based baseline testing as a means of documenting athletes’ levels of neurocognitive performance. Legislation is currently in place in 14 states in the United States requiring school districts to implement some type of concussion program, which may range from basic information dissemination to establishing policies and guidelines for concussion management. For example, the New Jersey State Interscholastic Athletic Association in its recent policy statement put forth formal recommendations to provide printed literature and annual concussion training to school personnel, student athletes, and parents. Ultimately, schools and athletic trainers are being called upon to protect student athletes from the dangers of concussion by adapting concussion testing programs that have been long-standing in professional sports.

The use of baseline testing for comparison with postconcussion test performance was first introduced in 1987 and has since been widely implemented in schools, universities, and professional sports organizations. The rationale behind this paradigm is to assist sports medicine personnel in making return-to-play decisions by comparing an athlete’s normal preconcussion neurocognitive functioning to his or her neurocognitive functioning after a concussion to help determine when recovery has been reached. The use of preseason baseline testing has been endorsed by sports concussion experts. Traditional paper-based tests have been deemed cost-inefficient and time-consuming, as they require many staff members for multiple test administrations,
but remain a valid means of assessment. In contrast, computer-based measurements can be completed with little supervision, even in the absence of a neuropsychologist.\textsuperscript{5,8}

Given that baseline testing is typically conducted in a computer laboratory or group setting in which a number of athletes can be tested simultaneously, the benefits of this time and cost-efficiency are shared by the organizations and personnel administering the tests.

However, group testing has received criticism and scrutiny. Baseline testing assesses cognitive skills such as attention, concentration, and memory. These tasks may be affected by distractions while testing, which could be expected to occur when a group of students, or a team of athletes, are present in the same room at the same time. Unfortunately, this paradigm shift, from individually administered paper-based testing to large-scale group-administered computerized testing, has not taken into consideration the effects of the testing environment on test performance, which can affect the validity of the tests used. Furthermore, available research regarding how the various consequences of this "shift" may affect test validity is inadequate.\textsuperscript{13,20}

Extraneous factors might be introduced when administering computer-based testing in groups. In any testing situation, environmental conditions, distractions, sensory deficits, and past or current psychological state of the test taker can influence the test taker's performance.\textsuperscript{17} However, it has been thought that the within-subjects nature of comparing an individual's baseline performance to his or her own postconcussion performance can control for these possible extraneous factors.\textsuperscript{17} But comparing a postconcussion performance, which was obtained in an individual test setting, to a baseline performance that was obtained in a group setting may be problematic because of the differences in test administration that may remove the control of extraneous factors and introduce variability.

To date, there has been little research comparing group and individual computerized test administration settings and effects on the cognitive performance. Early research identified no differences between individual and group test settings on simple mathematical and verbal tasks,\textsuperscript{7} as well as on a short-term memory task.\textsuperscript{5} In fact, social facilitation research,\textsuperscript{20} on the tendency to be more motivated to complete a task in the presence of others who are doing the same thing, generally supports increased performance when dealing with simple and easily learned tasks.\textsuperscript{2} However, researchers have found that participants are more distracted while working among an audience, and performance on complex tasks may be consequentially impaired.\textsuperscript{2} In addition, in a group setting, test takers may be inclined to conform to salient standards of the group, rather than task demand cues, and can therefore be influenced by the various distractions of a group test setting.\textsuperscript{4} In this context, students in smaller, more intimate test settings were found to perform better on stressful tasks than did students who completed the tests in large lecture halls.\textsuperscript{3} Similarly, it has been found that group-administered intelligence tests produce less stable IQ (intelligence quotient) scores than tests administered individually.\textsuperscript{22}

Suggestions for controlling testing conditions have been documented for decades, as test developers have recommended administration in quiet conditions with minimal distraction.\textsuperscript{21} Authors of intelligence tests\textsuperscript{28} requested that their tests be administered in settings that are quiet and disciplined. Nearly a century ago, it was suggested that voluntary concentration of the test taker is essential during cognitive testing, making it necessary to ensure for minimal environmental disturbance to produce a high-quality assessment.\textsuperscript{10} More recently, auditory\textsuperscript{29} and visual\textsuperscript{9} distractions have been linked with decreased task performance, and perceptual distraction (eg, background noise) has been linked to reduced cortical information processing and increased cerebral energy consumption during test-taking.\textsuperscript{27} Despite these findings, the environmental effects of individual versus group administration of neurocognitive baseline tests have yet to be investigated.

The purpose of this study was to evaluate differences in performance between high school athletes completing preseason, baseline neurocognitive testing in groups as compared with high school athletes completing baseline neurocognitive testing in an individual setting.

**METHODS**

**Participants**

Participants were high school–aged athletes who completed baseline neurocognitive assessments using the online version of ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing) during the 2007-2008 through 2009-2010 academic/athletic seasons. All baseline assessments were conducted through a relationship with a hospital-based neuropsychology practice. High schools were provided with licensed athletic trainers and/or concussion testing materials through a community outreach program with a goal of improved concussion education and increased participation in concussion testing and management. For those schools/teams choosing not to participate in the program, as well as for students in league sports in the community, offers were made to parents/guardians to obtain in-clinic assessments. Participants were retrospectively assigned to 2 independent groups: "group" setting or "individual" setting.

The "group setting" group consisted of 164 high school athletes (mean age, 15.7 years; standard deviation [SD], 1.15) who completed baseline neurocognitive testing in a group setting, in their high school computer laboratory, supervised by either a member of the school medical staff (ie, nurse or assistant nurse) or a licensed athletic trainer. These athletes attended 1 of 15 schools choosing to enter into a cooperative agreement with the hospital-based outpatient neuropsychology clinic. Testing of athletes was not mandated, and parents were able to refuse participation in the testing program. Of note, participants in the "group setting" group were randomly sampled from a larger sample of over 2300 high school athletes who completed baseline testing at their high schools in groups (see Procedures section below for more detail).
The “individual setting” group was 167 high school athletes (mean age, 15.2 years; SD, 1.31) who completed baseline neurocognitive testing in an individual setting, in the neuropsychological clinic. All athletes in this group were accompanied by a parent or guardian and were oriented to the test by a clinical neuropsychologist, who also provided feedback following the assessment (see Procedures section below for more detail).

The online version of the ImPACT test identifies potentially invalid baseline test results, based on the presence of outliers on select composite scores and subscales (ie, Impulse Control score >30), and these baseline results are “flagged” with the symbol “++.” Fifteen athletes (4.5%) were removed from the analyses because of suspected invalid baseline performance. The resulting samples were, thus, composed of 150 athletes in the group setting and 166 athletes in the individual setting groups.

Materials

All participants completed the online version of ImPACT. The ImPACT test was introduced in 2000 as a personal computer–based “desktop” version,11 and the online, cross-platform, Web-based version was introduced in 2008. The ImPACT test battery consists of 6 neuropsychologic tests, each designed to target different aspects of cognitive functioning including attention, memory, processing speed, and reaction time. From these 6 tests, 4 separate composite scores are generated: verbal memory, visual memory, motor speed, and reaction time. More thorough descriptions of ImPACT and its psychometric properties are available in the literature.12,16,23 The Post-Concussion Symptom Scale (PCSS)15 is also utilized in the ImPACT test battery to document and track concussion symptoms. The online version of ImPACT is identical to the desktop version, with the exception of the use of a keyboard input for a choice-reaction time task (“Xs and Os”) to reduce left-right confusion and subsequent user error.14

Analyses

Chi-square analyses were conducted to identify between-group differences in sex, attention deficit disorder/learning disability (ADD/LD), and history of concussion, and 1-way analyses of variance (ANOVAs) were conducted to identify between-group differences in age. Despite random sampling, the “individual” group was significantly younger ($F[3,1314] = 14.8; P = .001). As a result, assumptions for use of age as a covariate in between-groups analyses were evaluated.26 Age was significantly correlated with motor processing speed and reaction-time scores, but not with verbal or visual memory or total symptom scores. As well, homogeneity of regression slopes was observed for motor processing speed and reaction-time scores, so age was used as a covariate in analyses of covariance for these 2 composite scores. Univariate ANOVAs were conducted to identify between-group differences for the verbal and visual memory composite scores and the total symptom scores. Bonferroni correction (for 5 analyses) adjusted the alpha level to $P < .01 to decrease the

### TABLE 1

<table>
<thead>
<tr>
<th>Group Demographics&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Individual</th>
<th>Group</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>102 (61%)</td>
<td>115 (70%)</td>
<td>217 (66%)</td>
</tr>
<tr>
<td>Female</td>
<td>64 (39%)</td>
<td>50 (30%)</td>
<td>114 (34%)</td>
</tr>
<tr>
<td><strong>ADD/LD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>128 (85.3%)</td>
<td>105 (86.1%)</td>
<td>233 (85.7%)</td>
</tr>
<tr>
<td>1</td>
<td>17 (11.3%)</td>
<td>12 (9.8%)</td>
<td>29 (10.7%)</td>
</tr>
<tr>
<td>2+</td>
<td>5 (3.3%)</td>
<td>8 (4.1%)</td>
<td>10 (3.7%)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>15.2 (1.30)</td>
<td>15.7 (1.15)</td>
<td>15.42 (1.26)</td>
</tr>
</tbody>
</table>

<sup>a</sup>ADD/LD, attention deficit disorder; NR, not reported.

<sup>b</sup>As reported by athletes while completing the demographic portion of ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing).

rigorous controls, this retrospective study represents the current status of baseline testing in high schools.
TABLE 2
Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) Scores (Means and Standard Deviation) by Group

<table>
<thead>
<tr>
<th>Measure</th>
<th>Individual</th>
<th>Group</th>
<th>(F(1,314))</th>
<th>Significance (F)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal memory</td>
<td>86.53 (9.5)</td>
<td>83.37 (9.3)</td>
<td>8.95</td>
<td>.003</td>
<td>-.34</td>
</tr>
<tr>
<td>Visual memory</td>
<td>76.73 (12.9)</td>
<td>71.57 (12.4)</td>
<td>13.14</td>
<td>.0001</td>
<td>-.41</td>
</tr>
<tr>
<td>Motor processing speed(a)</td>
<td>38.35 (7.1)</td>
<td>35.58 (6.3)</td>
<td>13.52</td>
<td>.0001</td>
<td>-.42</td>
</tr>
<tr>
<td>Reaction time(a)</td>
<td>.577 (.08)</td>
<td>.607 (.08)</td>
<td>11.59</td>
<td>.001</td>
<td>.38</td>
</tr>
<tr>
<td>Symptom total</td>
<td>4.36 (9.4)</td>
<td>6.07 (9.3)</td>
<td>2.63</td>
<td>.11</td>
<td>.18</td>
</tr>
</tbody>
</table>

\(\text{Numbers in italics represent corrected, estimated marginal means, after covarying for age.}\)

TABLE 3
Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) Scores (Means and Standard Deviation) by Group, Excluding Athletes With Attention Deficit Disorder/Learning Disability and/or History of Concussion

<table>
<thead>
<tr>
<th>Measure</th>
<th>Individual</th>
<th>Group</th>
<th>(F(1,246))</th>
<th>Significance (F)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal memory</td>
<td>86.83 (9.5)</td>
<td>82.93 (9.3)</td>
<td>10.58</td>
<td>.001</td>
<td>-.41</td>
</tr>
<tr>
<td>Visual memory</td>
<td>78.07 (12.6)</td>
<td>71.02 (12.6)</td>
<td>19.36</td>
<td>.0001</td>
<td>-.56</td>
</tr>
<tr>
<td>Motor processing speed(a)</td>
<td>38.78 (7.0)</td>
<td>35.49 (6.4)</td>
<td>21.09</td>
<td>.0001</td>
<td>-.49</td>
</tr>
<tr>
<td>Reaction time(a)</td>
<td>.577 (.07)</td>
<td>.611 (.08)</td>
<td>17.33</td>
<td>.0001</td>
<td>.48</td>
</tr>
<tr>
<td>Symptom total</td>
<td>4.10 (9.7)</td>
<td>5.92 (9.5)</td>
<td>2.22</td>
<td>.14</td>
<td>.19</td>
</tr>
</tbody>
</table>

\(\text{Numbers in italics represent corrected, estimated marginal means, after covarying for age.}\)

The likelihood of type I error. Analyses were conducted using SPSS statistical software, version 16 (SPSS Inc, Chicago, Illinois).\(^{25}\)

RESULTS

The group distributions showed similar likelihood for sex \((\chi^2[1] = 1.48; P = .22)\), diagnosis of ADD/LD \((\chi^2[1] = 0.96; P = .33)\), and history of concussion \((\chi^2[1] = 0.25; P = .88)\) (Table 1). Individuals completing ImPACT in a group setting were significantly more likely to obtain invalid baseline results \((\chi^2[1] = 12.1; P = .001)\); only 1 athlete \((0.3\%)\) from the “individual setting” group obtained invalid baseline results, as compared with 14 athletes from the “group setting” group \((8.5\%)\).

Analyses of variance revealed a significant effect of test setting \((\text{individual vs group})\) on baseline neurocognitive performance, as measured by verbal memory \((F(1,314) = 8.95; P = .003; d = -.34)\) and visual memory \((F(1,314) = 13.14; P = .001; d = -.41)\). Even after covarying for between-groups differences in age, test setting had a significant effect on motor processing speed \((F(1,313) = 19.95; P = .001; d = -.42)\) and reaction-time scores \((F(1,313) = 15.86; P = .001; d = .38)\). No significant between-groups differences were observed on total symptom scores \((F(1,314) = 2.63; P = .11; d = .18)\). Means and standard deviations are provided in Table 2.

Excluding athletes with a history of concussion and/or history of ADD/LD \((N = 74)\) yielded a sample of 248 athletes. The ANOVAs revealed a significant effect of test setting \((\text{individual vs group})\) on baseline neurocognitive performance, as measured by verbal memory \((F(1,246) = 10.58; P = .001; d = -.41)\) and visual memory \((F(1,246) = 19.36; P = .001; d = -.56)\). Even after covarying for between-groups differences in age, test setting had a significant effect on motor processing speed \((F(1,246) = 21.09; P = .001; d = -.49)\) and reaction-time scores \((F(1,246) = 17.28; P = .001; d = .48)\). No significant between-groups differences were observed on total symptom scores \((F(1,246) = 2.22; P = .14; d = .19)\). Means and standard deviations are summarized in Table 3.

DISCUSSION

This study is the first to document decreased performance on baseline neurocognitive tests for athletes tested in a group environment, as compared with those tested in an individual setting. Athletes completing testing in a group setting scored significantly lower on all neurocognitive indices, compared with athletes completing testing in an individual setting, and there was greater incidence of invalid test performances in those who underwent group administration. It is important to note that even after excluding athletes with a history of concussion and/or ADD/LD, groups showed no significant differences in sex, and similar between-groups differences were noted on neurocognitive testing.

Baseline neurocognitive tests are routinely administered in groups, within high schools, colleges, and
universities, as a cost- and time-efficient way of making such services available to the greatest number of students. Scholastic aptitude and achievement tests, such as the Scholastic Aptitude Tests, are also administered in this manner. However, there are salient differences between achievement testing and neurocognitive baseline testing. First, students are well familiar with the parameters and style of achievement testing, with predictable, repetitive, multiple-choice formats that have been consistently presented year after year. In fact, students can even prepare for such tests with practice examinations, instructional computer programs, and test-taking classes to optimize performance. Second, neurocognitive tests measure skill sets that are quite vulnerable to distractions in the test environment. Typically and historically, neurocognitive testing should occur in a quiet, standardized, individualized setting. For example, neuropsychologists do not administer the Halstead Reitan Neuropsychological Test Battery, Conner's Continuous Performance Test, or the Wechsler Memory Scale (examples of neurocognitive tests that measure attention, concentration, memory, and other skills) in a group environment. Third, for those students who experience diagnosed attentional or learning disorders, test accommodations, such as extended time, breaks, and a separate room for testing, can be provided during school achievement testing. For baseline testing, providing extra time or breaks in the middle of the test would not only be impractical, but would actually invalidate the test performance as reaction time and speed are part of what is being measured.

The value of neurocognitive testing in the identification and management of concussion cannot be underestimated. The use of preseason baseline testing, as an additional component to aid in return-to-play decisions, has contributed greatly to the clinician's ability to make data-based judgments. However, the ability of any neurocognitive test instrument to provide valid and reliable data may, in part, be dependent upon the knowledge of the person administering/supervising the test and upon the proper test conditions. In the case of baseline testing, it is important for the person administering or supervising the test process to be knowledgeable about how to control the test environment, as well as how to screen for invalid test results.

Based on the current results, standardization and control of the test environment may be especially important during baseline testing because postconcussion testing is typically not performed in a group setting. In this context, postconcussion testing is less subject to the distractions and interruptions of baseline testing in the group format. However, because the purpose of baseline testing is for comparison with postconcussion test results, clinicians should ensure that they are accurately capturing the athlete's best and most consistent test performances both before and after a concussion. Paradoxically, decreased performance on tests administered in groups might compare favorably with postconcussion test performance, thus contributing to an earlier (although risky) return to play. While neurocognitive tests are not intended to be a lone diagnostic measure, the setting in which baseline data are obtained may be a relevant factor when assessing changes in neurocognitive function following a subsequent concussive injury.

To help reduce baseline test variability, we recommend due diligence of test administrators to (1) determine that athletes understand the purpose and nature of baseline testing; (2) ensure that athletes understand the test instructions; (3) encourage good effort on the part of the athlete and monitor test data for invalid results; and (4) reduce and control for distractions in the test environment. Specific ways to control the testing environment and experience may include (1) seating that is comfortably spaced, where athletes should not be directly next to or across from each other; (2) removal of extraneous sounds and interruptions, and use of sound proofing and/or white noise machines; (3) lighting that does not produce a glare on the computer screen; (3) clean and functional computer mice that are able to move with ease; (4) a test administrator who is present at all times; (5) clear group instructions and rules provided before testing regarding communication during the testing (raising hands if there is a computer glitch), with an opportunity for "Q and A" and bathroom breaks before starting; (6) identifying and removing any athletes who are overly talkative, not taking the test seriously, and/or are engaging in horseplay, before or during a group testing, and then testing them later individually; (7) testing when fatigue is not an issue, such as early in the day and not immediately following exercise, practice, or a game.

Consultation with a trained neurocognitive specialist when invalid baseline testing is evident, or if it persists despite the athlete's best effort, is advantageous and can explain confusing test results. For example, athletes who have been diagnosed with attention deficit hyperactivity disorder or a learning disorder may produce variable test results. In addition, conditions of longstanding right-left confusion or color blindness may significantly affect ImPACT results so that testing may appear invalid when in truth it is not.

This study is not without its limitations. While data were collected in independent settings, group assignment and methodologic controls were conducted retrospectively. As such, the experimental design is cross-sectional, without a priori controls or measurements for all demographic variables that could influence cognitive performance, such as intellectual ability or home environment that supports academic success. Furthermore, as a retrospective study, there was no control of test administrator assignment, a script of test instructions, or designated seating arrangements. As such, these limitations decrease the generalizability of the findings. While the purpose of the study was to illuminate performance differences in a group versus individual setting, it is possible that those athletes completing testing in an individual setting may have had greater motivation to perform well, especially because they knew beforehand that they would be receiving feedback. Such knowledge may have facilitated an expectancy effect that could have contributed to the group differences. For these athletes, a parent/guardian scheduled the testing, accompanied the athlete to the appointment, and was present for posttesting feedback.
With respect to group administration, athletes may have been less inclined to ask for clarification, help, or explanation in the presence of their peers. In addition, while the on-screen instructions remained constant, schools were not provided with a script for standardized instructions or introductory comments. In the absence of standardized procedures for test administration, it is not possible to determine the extent to which self-control, minimizing distractions, and courtesy to others actually took place or was encouraged. Importantly, future research is needed to systematically identify those specific variables that affect baseline test performance and validity, such as test administrator instructions and demeanor, seating arrangements, noise, effort, prior expectations, exercise before testing, and others. Nevertheless, this study captured the consequences of naturally occurring group neurocognitive testing and its decreased neurocognitive performance for high school athletes; whether the results can be generalized to collegiate or professional athletes is unknown.

The purpose of the present study is not to decrease the importance or utility of baseline neurocognitive testing or of any particular neurocognitive test. Rather, this study serves to educate those who administer or supervise baseline and concussion testing programs to be knowledgeable and sensitive to environmental conditions of testing. If we are to provide computerized testing programs in our schools to protect the brains of high school–aged athletes (and younger athletes), then it is imperative that schools be sure the programs are administered in the most effective manner.

REFERENCES

22. Roszkowski MJ. Stability of IQs from group-administered tests: some further data. Psychological Reports. 1984;54(2):482-482.