Introduction

Understanding and creating the conditions and environment where performance optimization can occur is one goal of most athletes, coaches, and leaders [2]. While it is difficult to isolate variables and understand the impact each can have on performance, research that does so is important for advancing the understanding of the human body and its limits. One variable known to affect decision-making and performance is sleep. As a necessary bodily function, many sleep studies compare sleep duration to overall performance [2]. In a study conducted to determine the influence of sleep quality and duration on school performance, researchers found that sleep quality and sleep duration had significant impacts on school performance [3]. While studies like this one are important in determining the overall effect of sleep on performance, studies that look at metrics of performance, such as reaction time, can be more useful in understanding the specific effects sleep duration and sleep deprivation can have. Taheri et al. [2] sought to understand the effect of sleep deprivation on choice reaction time and anaerobic power. The researchers took baseline measurements from 18 collegiate athletes and had each perform a Wingate test on the cycle ergometer to test anaerobic power. In addition, subjects also performed a manual, two-choice reaction task on the computer to test choice reaction time. Both tests were performed after a normal night’s sleep and again after 24-hour sleep deprivation. The results show that the peak power of the subjects after sleep deprivation (8.3 ± 1.6 w.kg⁻¹) was not significantly changed from the baseline (7.9 ± 1.3 w.kg⁻¹; p=0.3). Statistical analysis of the choice reaction time results found that the mean choice reaction time of subjects exposed to sleep deprivation (281.65 ± 31 ms) was significantly slower than the baseline mean reaction time (244 ± 39 ms; p=0.003) [2].

Similar findings came from another study that examined the effect of sleep deprivation on cognitive and physical performance in college students [4]. This study required subjects to perform a cognitive function test on a mobile application that focused on working memory; a physical function test consisting of sub maximal cardiopulmonary exercise testing, and reaction time tests including the ruler drop test and computerized assessments. One group was allowed to get uninterrupted sleep while the other was instructed to refrain from sleeping and using any stimulants the night before. The researchers found that sleep deprivation had no significant
were instructed to press a button when they saw a color change on the computer screen. The results found that nurses working the day shift had a significantly quicker reaction time (0.33±0.14s) compared to those working the night shift (0.54±0.15s; p<0.0001).

In each test, nurses working the night shift performed worse than those working the day shift. The difference in cognitive functioning and reaction time indicates that sleep deprivation and the disturbance of circadian rhythm can negatively impact mental and cognitive performance [7].

Sleep studies often target sleep deprivation and extreme sleep loss; however, a study conducted by Cote et al. [8] aimed to understand the impact of lesser degrees of sleep restriction on brain functioning. One aspect of brain functioning tested was reaction time. For the study, participants were instructed to complete performance assessment batteries (PABs) including a reaction time task that required participants to press “0” on the keyboard whenever they heard a 70-dB tone. Following completion of the PABs, participants began a 96-hour protocol, during which time all participants remained in a lab under constant supervision by research assistants. On the first night, all participants were permitted 8 hours to go to bed (23:00 to 07:00 hours). On the second and third night, participants were randomly placed into one of three groups. One group was allowed 3 hours in bed, another 5 hours, and the third group was allowed 8 hours. On night 4, all participants were allowed 8 hours in bed. The results of this study show that mean reaction time in the 8 hour group remained stable across the 960-hour study. The 5 hour group showed slower mean reaction time during night 1 (283.31±44.51ms) and night 2 (291.99±49.26ms) compared to the baseline mean reaction time (259.34±26.57ms; p<0.05). The 3 hour group was slower than the baseline test (253.40±23.18ms) on night 1 (284.18±55.32ms) and on night 2 (301.30±64.62ms; p<0.05) [8].

There are many different ways to test reaction time. In a study conducted by Khitrov et al. [9], a device labelled the Psychomotor Vigilance Task-192, (PVT-192), is described as the “gold standard for simple visual reaction time testing.” The device consists of a left and right button and an LED dot-matrix display that shows a four-digit millisecond counter. The study’s purpose was to determine the extent to which the reaction time from the PVT-192 matched the reaction times found by a device the researchers developed called the PC-PVT (personal computer-psychomotor vigilance task). To do so, reaction time data was collected from both the PVT-192 and various computers using the PC-PVT software. The percent error was then calculated for each device. The results of the study showed that percent error when using a gaming mouse in conjunction with the PC-PVT was 3% compared to a percent error of 1% for the PVT-192. Based on this information, the researchers determined that the PC-PVT had only small differences in quality of data compared to the PVT-192 and can be used to conduct simple reaction time testing [9].

Another device developed to test reaction time is the Dynavision™ D2 Visuomotor Training Device (D2). The D2 is a light-training reaction device, developed to train sensory motor integration through the visual system. It consists of a board (4x4ft.) that can be used to test reaction time data.
raised or lowered relative to the height of the participant. The board contains 64 target buttons arranged into five concentric circles that can be illuminated to serve as a stimulus for the participant, and contains an LCD display above the innermost ring of target buttons. The LCD display provides a 5 second visual countdown to the start of a test [10]. A study was conducted to determine the reliability of the D2 as a reaction time test using intra class correlation coefficients (ICC). 42 participants reported to a Human Performance Lab six times with no less than 48 hours between sessions. During each session, subjects completed three consecutive assessments of increasing difficulty. The first was a choice reaction time test (CRT), the second a reactive test, and the third a reactive test with cognitive stress. The analysis of the reaction time data using repeated measures ANOVA and ICC indicate that the visual reaction time has strong reliability (ICC=0.84) and motor reaction time (ICC=0.63) has moderate reliability. Based on these results, it was determined that the D2 is a reliable test of reaction time [10].

While advances in software make computer testing ideal and most accurate, this method may not always be practical or available. The Ruler Drop Method (RDM) is another method used to test reaction time that is field-expedient and widely accepted as a valid method for testing reaction time. Aranha et al. [11] performed a study that compared the RDM to a mobile-based software application. After comparing the reaction times, Aranha et al. [11] found Spearman’s $\rho$ (rho 0.54; $p$=0.031). This indicates that the RDM is a moderate to good method for determining reaction time.

Another study conducted to assess the reliability and validity of the Ruler Drop Method utilized a clinical reaction time apparatus designed to emulate a ruler. The apparatus used was a 1.3m dowel rod with a weighted disk at the bottom of the rod intended to keep the rod vertical [12]. During this study, 65 participants were given the clinical reaction time with the rod ($RT_{clin}$) and a computerized reaction time test ($RT_{comp}$). $RT_{clin}$ required participants to rest their dominant forearm on a flat desk with their wrist over the edge of the desk and their hand open. The apparatus was suspended vertically and released at random time intervals, at which point the participant caught it as quickly as they could [12]. Using the equation $d = \frac{1}{2} gt^2$, the distance the apparatus travelled was converted to time in milliseconds. For $RT_{comp}$ participants sat at a computer with their dominant hand over the keyboard and pressed the spacebar whenever the black circle on a white background changed to a black “X”. The results show that $RT_{clin}$ had excellent test-retest reliability (ICC=0.860; $p=0.004$) and interrater reliability (ICC=0.915; $p=0.001$). There was also a significant positive correlation between the mean reaction time found in both tests [12].

Similar findings came from a study conducted on National Collegiate Athletic Association Division I athletes. Eckner et al. [13] used a stick with a weighted rubber disk as a clinical evaluation of reaction time ($RT_{clin}$) and compared it to a computerized reaction time test ($RT_{comp}$). Data from each test was collected 1 year apart during preseason physical examinations. The ICC estimates from season 1 to season 2 were 0.645 for $RT_{clin}$ and 0.512 for $RT_{comp}$. The similarity in ICC values between $RT_{clin}$ and $RT_{comp}$ indicate that $RT_{clin}$ compares favourably with that of $RT_{comp}$ supporting $RT_{data}$ potential use as part of a multifaceted concussion assessment battery [13].

Though the previous study accepts the RDT as a part of a multifaceted concussion protocol, it does not address possible learning that occurs from multiple trials of the RDT. Del Rossi, Malaguti et al. [14] performed a descriptive laboratory study designed to determine if the RDT is susceptible to practice effects after continuous administration. The design used a 60cm long measuring stick dropped between the fingers of the dominant hand, dropped at random intervals between 1s and 5s to prevent anticipatory responses. The distance the ruler dropped was converted to time using $d = \frac{1}{2} gt^2$. Each participant completed 10 trials on 10 different sessions. The 3 fastest and slowest times recorded for each session were eliminated as potential outliers. Mean reaction time was calculated and a repeated measures ANOVA performed. Statistical analysis showed that simple reaction time decreased after repeated assessments, the greatest reduction occurring between the first session (264.9±17.1 ms) and second session (257.7±18.2 ms). This indicates that the RDT is susceptible to practice effects [14].

Besides practice effects, another possible variable when measuring reaction time is subject motivation. Eckner et al. [15] performed a cross-sectional, observational study designed to investigate the influence of performance feedback and motivation during two tests of simple reaction time. One test, $RT_{clin}$ used the RTD to test reaction time while $RT_{comp}$ used a computerized RT testing with ($RT_{comp}$FB) and without ($RT_{comp}$FB) performance feedback. Repeated measures ANOVA analyzed the means and standard deviations of each test and Pearson correlation coefficients were calculated to assess the relationships between tests. The results of these analyses found that mean R and RT variability differed significantly for the 3 tests ($RT_{clin}$=234±28 ms; $RT_{comp}$FB=301±45 ms; $RT_{comp}$NoFB=327±52 ms; $p=0.0001$). They also found that participant motivation levels differed significantly across the three tests ($p<0.0001$). Participants found $RT_{clin}$ more highly motivating than the two computerized tests [15].

Collecting sleep data involves more complex measures. The current “gold standard” for sleep/wake identification is polysomnography (PSG), a method which uses electroencephalography (EEG) to record brain activity [16]. While PSG provides the most accurate data, it requires a laboratory and a trained professional. In an effort to provide convenient and cost-effective sleep data, actigraphy has been used as an alternative to PSG. Actigraph devices are movement detectors that rely on accelerometers. There are many different types and brands of commercially available actigraphy devices. In a study by Rupp & Balkin [16], data from two actigraphy devices, the Motion logger Watch and the Acti watch, was compared to PSG data to compare the two devices. During the experiment, 29 participants were first given 8 hours in bed from 23:00 to 07:00 and then stayed awake for 36 hours straight. Each participant wore both the Motion logger Watch and the Acti watch on the non-dominant wrist. The Motion logger Watch and the Acti watch were compared to PSG, a method which uses electroencephalography (EEG) to record brain activity [16]. While PSG provides the most accurate data, it requires a laboratory and a trained professional. In an effort to provide convenient and cost-effective sleep data, actigraphy has been used as an alternative to PSG. Actigraph devices are movement detectors that rely on accelerometers. There are many different types and brands of commercially available actigraphy devices. In a study by Rupp & Balkin [16], data from two actigraphy devices, the Motion logger Watch and the Acti watch, was compared to PSG data to compare the two devices. During the experiment, 29 participants were first given 8 hours in bed from 23:00 to 07:00 and then stayed awake for 36 hours straight. Each participant wore both the Motion logger Watch and the Acti watch on the non-dominant wrist while simultaneously PSG data was collected. To analyze the data, two analyses were performed: epoch-by-epoch agreement with discriminability measures (d’) and sleep pareme-
ter concordance. Repeated measures ANOVAs were performed for all variables. The results showed a significantly (p<0.001) higher sensitivity (sleep identification) (96.2±3.6 %), specificity (wake detection) (63.6±28.1 %), and overall agreement (93.6±4.0 %) with PSG for the Motion Logger Watch compared to the Actiwatch whose sensitivity was 92.2±2.9 %, specificity was 57.6±23.6 %, and overall agreement was 89.6±3.5 %.

While PSG and actigraphy are two of the most accurate objective measures of sleep duration and quality, access to these devices is not always available or feasible. In such cases, subjective measures of sleep can offer information about subjects' sleep quality and quantity. The Pittsburgh Sleep Quality Index (PSQI) is a self-reported survey often used to quantify sleep. It consists of seven component scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, sleep medication, and daytime dysfunction [17]. A study was conducted to test the reliability and validity of the PSQI along with the Epworth Sleepiness Scale (ESS), a self-reported survey used to identify excessive daytime sleepiness. Over the course of a month, 3,059 men were asked to complete a PSQI each night, an ESS each day, and wear wrist actigraphy for three or more 24-hour periods. 2,555 of these participants were excluded. Upon statistical analysis, the PSQI was found to have an adequate internal consistency at α=0.69. However, in this study, PSQI was found to be more strongly associated with subjective psychosocial and functional measures rather than actigraphic variables [18].

Another study conducted by Landry et al. [18] sought to understand the comparable between self-reported sleep in the PSQI and subjective measures of sleep recorded by actigraphy. To do so, sleep quality and quantity was objectively measured by the motion watch 8© (MW8) actigraphy system for 14 days. Researchers created a categorical composite score for the MW8 data that aligned with the PSQI categories of “good” and “poor” and “average”. Subjects recorded a subjective measurement of sleep quality using the PSQI for the month prior to the 14 day MW8 recordings. Upon awakening each morning during the 14 day assessment, subjects filled out a consensus sleep diary (CSD), another subjective measure of sleep quality. After analyzing the data using Pearson’s correlation, it was found that in correlation between MW8 duration data and PSQI sleep quality score was modestly correlated (r=0.32; p<0.01) [18].

Lauderdale et al. [19] also used actigraphy monitors and PQSI along with the ESS and Berlin Questionnaire. The intent was to compare the actigraphy objective data to the subjective data collected and determine the correlation between the two. Participants were mailed actigraphy monitors and the three questionnaires and asked to wear the monitors for three nights, preferably Wednesday-Friday. To compare the objective and subjective measures of sleep, a Pearson’s correlation was found. The results of this analysis show that on average, people who slept 5 hours self-reported 6.29 hours of sleep while those who slept 7 hours reported 7.31. The correlation found between the subjective and objective sleep measurements was 0.45, which is considered a moderate correlation [19].

While understanding the different methods of testing both sleep and reaction time are important, understanding the impact reaction time has on various performance metrics is the main purpose of determining sleep’s effect on reaction time. One area that can be severely affected by delayed reaction time is driving. Guo et al. [20] performed a study to analyze the relationship between driving fatigue, physiological signals and driver’s reaction time. To do so, subjects were asked to take a driving stimulation where they remained driving for up to 6 hours. Reaction time was tested throughout and a self-assessment of sleepiness recorded using the Stanford Sleepiness Scale. The results of the study show that the relationship between reaction time and driving fatigue had the largest correlation with accuracy of 86%, sensitivity of 87.5% and specificity of 85.52%. Overall, the results show that increased reaction time is correlated with driving fatigue [20].

References


