The Effects of Sleep Deprivation on Simulator Driving as Compared with Other Psychomotor Tests George

D Park
Systems Technology, Inc., Hawthorne, CA

Marie Claude Ware
Eastern Virginia Medical School, Norfolk

Jennifer F May
Eastern Virginia Medical School, Norfolk

Theodore J Rosenthal
Systems Technology, Inc., Hawthorne, CA

Matthew R Guibert
Systems Technology, Inc., Hawthorne, CA

See next page for additional authors

Follow this and additional works at: https://ir.uiowa.edu/drivingassessment

Park, D; Ware, Marie Claude; May, Jennifer F; Rosenthal, Theodore J; Guibert, Matthew R; and Allen, R Wade. The Effects of Sleep Deprivation on Simulator Driving as Compared with Other Psychomotor Tests George. In: Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, July 9-12, 2007, Stevenson, Washington. Iowa City, IA: Public Policy Center, University of Iowa, 2007: 257-264. https://doi.org/10.17077/drivingassessment.1246

Presenter Information

D Park, Marie Claude Ware, Jennifer F May, Theodore J Rosenthal, Matthew R Guibert, and R Wade Allen

THE EFFECTS OF SLEEP DEPRIVATION ON SIMULATOR DRIVING AS COMPARED WITH OTHER PSYCHOMOTOR TESTS

George D. Park,[1] J. Catesby Ware,[2] Jennifer F. May,[2]
Theodore J. Rosenthal,[1] Matthew R. Guibert,[1] and R. Wade Allen[1]
[1] Systems Technology, Inc., Hawthorne, CA
[2] Eastern Virginia Medical School, Norfolk, VA
E-mail: georgepark@systemstech.com

Summary: A dose-response sleep deprivation study was performed using 14 participants of normal health (7 males, 7 females, age 19-57 years). Participants completed 3 separate 24-hour visits where they were allowed to sleep for either 8, 4, or 0 hrs and then asked to complete a test battery that included three psychomotor test measures known to be sensitive to the effects of fatigue and extended wakefulness. The psychomotor performance measures included the standard deviation of lateral lane position (SDLP) for a driving simulator, the Critical Tracking Task (CTT), and the Psychomotor Vigilance Task (PVT). While all psychomotor test measures provided a significant main effect for the within-subjects variable sleep condition, subsequent pairwise comparisons between the three sleep conditions provided clear differences in the strengths of the psychomotor measures. SDLP measures provided the best discrimination between the sleep conditions when utilizing the full 60-min driving task time, but indicated a better sensitivity at lower levels of sleep loss. The CTT provided the contrary, with better sensitivity at higher levels of sleep loss. The PVT measures of mean reaction time and number of lapses and slowest 10% reaction time failed to discriminate between the different sleep conditions.

INTRODUCTION

A variety of psychomotor tests have been used in assessing impairments associated with fatigue and extended wakefulness. For example, the Psychomotor Vigilance Task (PVT) is a simple sustained-attention and reaction time test found to be sensitive to the effects of daytime fatigue or sleepiness and unaffected by learning or practice effects (Balkin et al., 2000). Methodologically reliable and relatively versatile, the PVT is a standard laboratory tool for the assessment of sustained performance in a variety of experimental conditions involving dose-response sleep studies (Loh, Lamond, Dorrian, Roach & Dawson, 2004).

While the PVT is example of a discrete psychomotor movement test, continuous psychomotor movement tests, particularly tracking tasks, have the potential to provide a even more sensitive measure of fatigue or motor impairment. According to Schmidt and Lee (1999), there are basically two kinds of tracking tasks: pursuit and compensatory. In pursuit tracking, the variations of the track pursued by a subject and the subject's own movements are both displayed. An example of pursuit tracking is steering a vehicle, where the roadway is the track and the vehicle provides the subject's movements.

To measure the pursuit tracking ability of a driver under the effects of fatigue or sleepiness, a common and simple performance measure is the vehicle lane variability (aka, lane drifting, swerving, or weaving), which can be defined as the standard deviation of vehicle lateral lane position (SDLP; O'Hanlon, 1984). As a driver's psychomotor performance becomes impaired, the driver's ability to maintian vehicle position in the road lane degrades and SDLP scores increase. High correlations have been found between sleepiness and SDLP performance (Contardi, Pizza, Sancisi, Mondini & Cirignotta, 2004). SDLP performance has also been found to have good test-retest reliability (Marcotte, Roberts, Rosenthal, Heaton & Bentley, 2003) and the ability to discriminate drivers with sleep disorders such as obstructive sleep apnea (Risser, Ware & Freeman, 2000).

In compensatory tracking, the variations of the track are combined with the subject's movements to provide a single displayed value, and the subject's goal is to maintain this value at some constant location. An example of compensatory tracking is an aircraft instrument such as the glide scope indicator, where the difference between the proper altitude and the actual altitude is displayed; when the dial is is in the middle of the screen, the pilot's altitude is correct. Compensatory tracking tasks are considered generally more difficult than pursuit tracking tasks, particularly if the behavior of the track is irregular and unpredictable (Schmidt & Lee, 1999).

The Critical Tracking Task (CTT) is example of a psychomotor test that measures the compensatory tracking ability of a subject. The task requires continuous hand-eye coordination in order to keep a laterally unstable moving screen object within a specified section of a computer screen. The longer a participant is able to maintain control, the higher the performance score obtained. Different types of user controls (e.g., trackball, joystick or steering wheel) can be configured to make the CTT more ecologically valid. Originally developed to measure the capability of pilots and astronauts (Jex, McDonnell & Phatak, 1966), the CTT has subsequently been applied to detecting driver alcohol impairment , truck driver fatigue, and worker impairment for fitness-for-duty (Allen, Stein & Miller, 1990).

The purpose of this paper is to compare the results of the PVT, the SDLP of the driving simulator, and the CTT psychomotor tests in relation to a dose-response sleep deprivation study. The overall objective of the experiment was to determine simple, objective and efficient measures of fitness-for-duty for sleep-deprived personnel. Participants were then given a test battery that included psychomotor, subjective, and physiological measures. The results of the psychomotor measures are presented here and discussed.

METHODS

Participants

A total of 14 participants of normal health (7 males, 7 females, age 19-57 years) participated in this sleep deprivation study. All participants were healthy young adults, screened for sleep disorders, stimulant/depressant medications and Epworth Sleepiness Scale (ESS) scores. If participants scored high on the ESS, were taking sleep-influencing medications, or had a sleep disorder, they were excluded from the study. Participants received $500 compensation.

Procedures

Training and baseline data was collected over a 2-week period for the CTT and driving simulator. After the completion of the training period, each participant completed 3 separate 24-hour visits where they were allowed to sleep for either 8, 4, or 0 hrs. In the morning, participants completed the test battery. Participants were run in pairs, experienced the same sleep condition at the same time, were allowed to interact with each other to keep themselves awake, and were also monitored by sleep technicians via video and audio to ensure wakefulness.

Psychomotor Test Measures

Psychomotor Vigilance Task (PVT). Under each sleep-deprivation condition, each participant completed a 10-min PVT session. The PVT involved a book-sized, hand-held device that has two response buttons and an LED four-digit numeric display. Subjects were instructed to press a response button as quickly as possible after the display appeared and began counting. Inter-stimulus interval varied randomly from 2- to 10-second increments. Each test session was generally 10-min long with anywhere from 120 to 140 stimuli instances. Scores for overall mean reaction time (ms) and number of lapses (response latency exceeding 500 ms) are provided here.

Driving Simulator. A fixed-platform, PC-based driving simulator consisting of force feedback steering, gas and brake pedals, a single 50 x 50 in. (127 x 127 cm) roadway projection display, surround sound, and a road vibration system was used. The seat, controls, video monitoring, and roadway display system were housed in a 5 x 10 ft (approximately 1.5 x 3 m) room covered with sound-attenuation carpeting. The simulator setup and placement of controls were similar to the specifications of an automatic transmission vehicle. The experimenter, computers, and the video–audio monitoring equipment were in an adjacent room.

A 10-min practice drive was administered during the training period, which consisted of a city scenario that included intersections, stoplights, traffic, and pedestrians that required participant responses. The test drive consisted of a simple, rural highway scenario for a drive time duration of 60 min at a speed of 55 mph (88.5 km/h) with long, wide curves and an oncoming vehicle every 10 min (Risser, Ware et al., 2000). This type of drive required participants to maintain vigilance to stay on the road, but required no other responses. The SDLP performance data of each participant was collected at six 10-min markers for each sleep condition.

Critical Tracking Task (CTT). The CTT display in this experiment was configured as a driving task, as illustrated in Figure 1, and controlled with a Microsoft Sidewinder Force Feedback PC steering wheel. The basic task of each participant was to use the steering wheel to keep the vehicle from touching the red bars located on the sides of the computer screen. The lateral instability rate of the vehicle was set to begin easy and then quickly increase linearly with time until the participant could not keep the vehicle within the green colored bar section. At that point, the instability rate would lessen to fine-tune a trial performance score. Inevitably, the vehicle becomes so unstable that the participant loses control, the vehicle touches either red bar and the trial finishes. Two training sessions were provided, approximately 15 trials each. During the testing sessions, each participant was given a total of 8 trials to provide an average CTT performance score. Average CTT test session duration was 4 to 5 min.
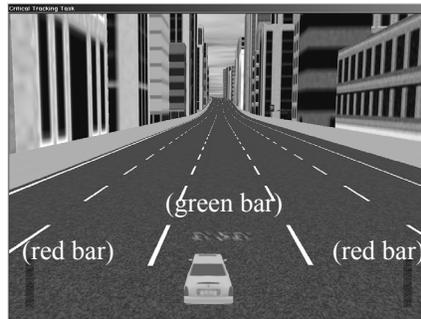
Figure 1. The Critical Tracking Task (CTT) screen display

RESULTS

Psychomotor Vigilance Task

Separate one-way within-subjects ANOVAs were performed for PVT mean reaction time scores, and number of lapses, $N = 14$, $\alpha = .05$. The within-subjects variable was sleep conditions (8hr, 4hr, and 0hr). For mean reaction time, a significant main effect was found for sleep condition, $F(1.25,19.47) = 5.79$, $p = .023$, $\eta^2 = .308$, when the Greenhouse-Geisser correction was used to adjust for sphericity violation. As shown in Figure 2a, mean reaction time scores and variability increased as participants experienced lesser hours of sleep, 8hr ($M = 270.26$, $SD = 39.25$), 4hr ($M = 298.01$, $SD = 69.09$) and 0hr ($M = 361.29$, $SD = 132.18$). However, Bonferroni pairwise comparisons only found a significant difference between the 8hr and 0hr sleep conditions ($p = .049$). For number of lapses, a significant main effect for sleep condition was found, $F(2,26) = 4.97$, $p = .015$, $\eta^2 = .277$. As shown in Figure 2b, mean number of lapses and variability also increased as participants experienced lesser hours of sleep, 8hr ($M = 2.21$, $SD = 2.94$), 4hr ($M = 4.79$, $SD = 7.77$), and 0hr ($M = 9.79$, $SD = 11.91$). Bonferroni pairwise comparisons failed to find any significant differences between any of the sleep conditions.
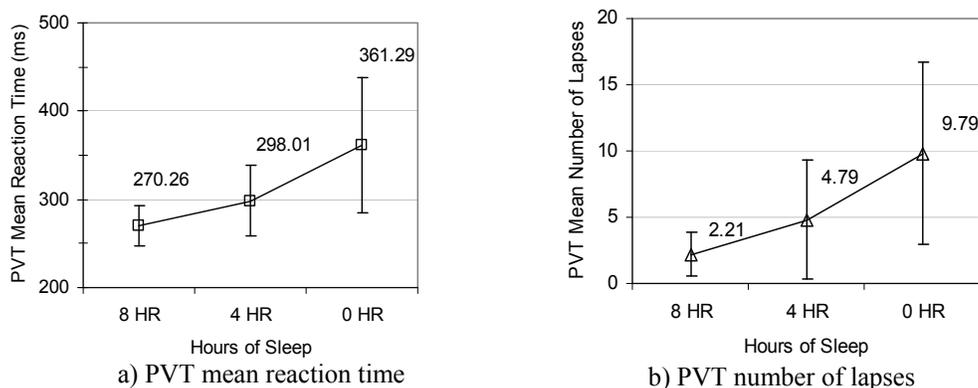

a) PVT mean reaction time

b) PVT number of lapses

Figure 2. Psychomotor vigilance task (PVT) across sleep conditions with 95% confidence intervals

Driving Simulator

A 3 x 6 within-subjects ANOVA was performed for the driving simulator performance measure SDLP, $N = 14$, $\alpha = .05$. The within-subject variables included sleep condition (8hr, 4hr, and 0hr) and driving time (10, 20, 30, 40, 50, 60min markers). A significant main effect was found for sleep condition $F(2,26) = 14.79$, $p < .001$, $\eta^2 = .532$, with mean SDLP scores increasing in size as participants experienced lesser hours of sleep (8hr: $M = 1.15$, $SD = .145$, 4hr: $M = 1.64$, $SD = .309$, and 0hr: $M = 2.33$, $SD = .490$). Bonferroni pairwise comparisons confirmed this trend with significant differences found between 8hr and 4hr ($p = .008$), 8hr and 0hr ($p < .001$), and 4hr and 0hr ($p = .019$) conditions. A significant main effect was also found for driving time, $F(2.44,3.45) = 10.42$, $p < .001$, $\eta^2 = .445$, when the Greenhouse-Geisser correction was used to adjust for sphericity violation. A significant linear, $F(1,13) = 20.01$, $p = .001$, $\eta^2 = .606$, and significant quadratic trend, $F(1,13) = 11.42$, $p = .005$, $\eta^2 = .468$, was also found. Figure 3a provides the SDLP means for each sleep condition across driving time.

To examine the utility of a 60-min simulator driving time for SDLP performance, 6 separate one-way within-subjects ANOVAs were performed for each of the 6 driving time markers, $N = 14$, $\alpha = .05$. The within-subjects variable was sleep condition. As expected, a significant main effect was found for sleep condition for each driving time marker, with all p values $< .002$. However, subsequent Bonferroni pairwise comparisons between the sleep conditions at each time marker were not consistent.

Significant differences were found between the 8hr and 0hr sleep conditions at all 6 driving time markers (all p values $< .005$), suggesting that participant SDLP performance was capable of discriminating participants when provided 8hrs of sleep as oppose to zero after only 10 min and throughout the 60-min driving time. Figure 3b provides the 8hr and 0hr sleep condition scores across driving time.

Significant differences were found between the 8hr and 4hr sleep conditions at driving time markers 10 to 50min (all p values $< .045$), but not at the 60min marker. Similar to the 0hr sleep condition, score variability for the 4hr sleep condition was considerably larger in magnitude compared to the 8hr sleep condition as shown in Figure 3c. Except for the 60min marker where performance in the 4hr sleep condition appeared to improve, there appeared to be little overlap in score variability throughout the first 50 min of the driving simulation test.

Significant differences were found between the 4hr and 0hr sleep conditions at only driving time markers 10min ($p = .032$) and 60min ($p = .012$). No differences were found between 4hr and 0hr sleep conditions for markers 20 to 50min. The overlap in score variability for those markers can be seen in Figure 3d.
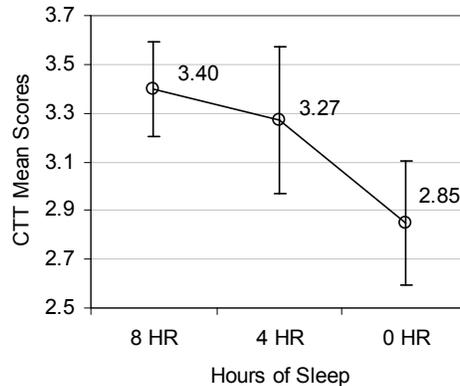
Critical Tracking Task

A one-way within-subjects ANOVA was performed for CTT performance scores, $N = 14$, $\alpha = .05$. The within-subjects variable was sleep conditions (8hr, 4hr, and 0hr). The dependent variable was the average CTT score from 8 testing trials. A significant main effect was found for sleep condition, $F(2,26) = 11.25$, $p < .001$, $\eta^2 = .464$, indicating that when participants received

fewer hours of sleep, decreases in CTT performance were found, 8hr (M = 3.40, SD = .333), 4hr (M = 3.27, SD = .526), and 0hr (M = 2.85, SD = .440). Bonferroni pairwise comparisons provided significant differences between 8hr and 0hr (p = .002) and 4hr and 0hr (p = .032) conditions. No difference was found, however, between the 8hr and 4hr condition. Figure 4 provides mean CTT scores for each sleep condition.

a) SDLP for each sleep conditions

b) SDLP for 8hr versus 0hr

c) SDLP for 8hr versus 4hr

d) SDLP for 4hr versus 0hr
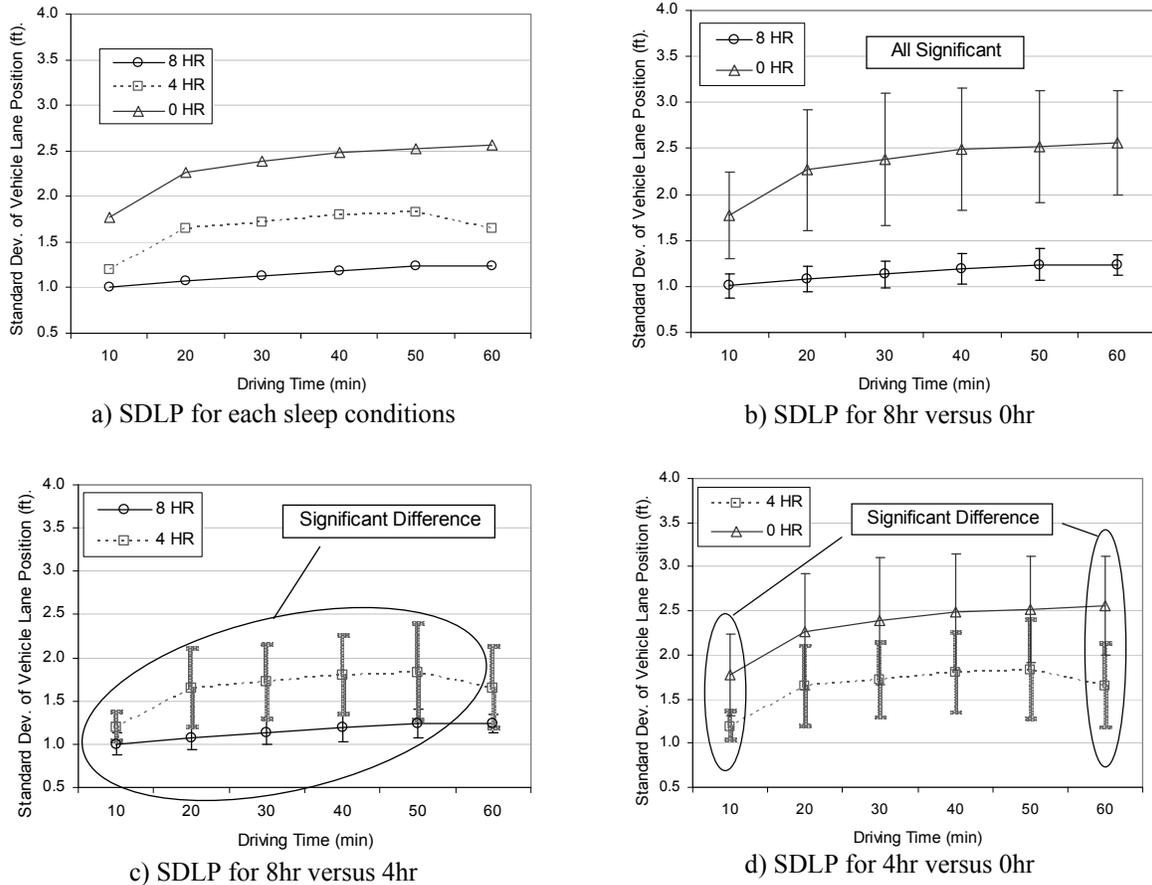
Figure 3. Driving simulator standard deviation of vehicle lane position (SDLP) for each sleep condition across driving time markers with 95% confidence intervals

Figure 4. Critical tracking task (CTT) mean scores across sleep conditions with 95% confidence intervals

DISCUSSION

In general, the performance measures from all three psychomotor test measures were able to discriminate participants when they were given 8 hrs of sleep versus 0 hrs of sleep. However, subsequent pairwise comparisons between the three sleep conditions provided clear differences in the strengths of the psychomotor measures. The PVT measures of mean reaction time and number of lapses provided the least sensitive measure of sleepiness. While main effects for sleep condition were found for both measures, mean reaction time was only able to discriminate between those participants with 8hrs of sleep versus total sleep deprivation. Number of lapses failed to discriminate between any of the sleep conditions.

When utilizing the full 60-min of driving task time, the driving simulator as a pursuit tracking task appeared to provide the most sensitive measure to sleep deprivation, with lane deviations doubling in size under the no-sleep condition and significant pairwise comparisons between all sleep conditions. However, subsequent analyses at each driving time marker revealed that from 20 to 50 min, SDLP performance was unable to discriminate the participants from the 4hr and 0hr sleep conditions, suggesting that SDLP performance is most discriminating at lower levels of sleep loss and not necessarily at higher levels of sleep loss.

Conversely, the CTT was the most time efficient fatigue test and discriminated performance between higher levels of sleep loss (4hr to 0hr), but not lower levels (8hr to 4 hr). The results suggest the driving simulator task, which can be described as monotonous and requiring sustained-vigilance, can be used to draw out fatigue differences in lower levels of sleep loss. Whereas the CTT task, which requires more difficult compensatory tracking movements, may be insensitive to lower levels of sleep loss (especially when subject motivation can still be a factor) but increasingly sensitive as sleep loss becomes more severe. Further studies of the CTT when modified to be "easier" in variability and unpredictability may provide a fitness-for-duty test that can be both efficient and as discriminating as a pursuit tracking driving simulator task.

A point of interest in the SDLP data was at the 60-min marker, where mean participant scores under the 4hr sleep condition appeared to actually improve. This unexpected score decrease most likely caused the non-significant difference between the 8 and 4hr condition, and the significant difference in the 4 and 0hr condition. This may have been due to any number of individual participant differences, how they cope/perform under different levels of fatigue/sleepiness, which were not addressed in this paper but have been found to be important for SDLP measures (Ingre, Akerstedt, Peters, Anund & Kecklund, 2006). Further analysis of the data with subjective sleepiness scores for individual differences is warranted here.

REFERENCES

Allen, R.W., Stein, A.C., & Miller, J.C. (1990). Performance testing as a determinant of fitness-for-duty. *SAE Aerospace Technology Conference & Exposition*, Long Beach, CA, Oct 1-4.

Balkin, T.J., Thome, D., Sing, H.C., Thomas, M.L., Redmond, D.P., Wesensten, N.J., et al. (2000). *Effects of sleep schedules on commercial motor vehicle driver performance* (DOT-MC-00-133). Washington, DC: Department of Transportation Federal Motor Carrier Safety Administration.

Contardi, S., Pizza, F., Sancisi, E., Mondini, S. & Cirignotta, F. (2004). Reliability of a driving simulation task for evaluation of sleepiness. *Brain Research Bulletin,* 63(5): 427-431.

Ingre, M., Akerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink durations: Examining individual differences. *Journal of Sleep Research, 15*(1): 47-53.

Jex, H.R., McDonnell, J.D., & Phatak, A.V. (1966). A "Critical" Tracking task for manual control research. *7th IEEE Symposium on Human Factors in Electronics (HFE)*, HFE-7, 138-145.

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-min duration. *Behavior Research Methods, Instruments, & Computers, 36*(2): 339-346.

Marcotte, T.D., Roberts, E., Rosenthal, T.J., Heaton, R.K., & Bentley, I.G. (2003). Test-retest reliability of standard deviation of lane position as assessed on a PC based driving simulator. *Proceedings of the 2nd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*,

O'Hanlon, J.F. (1984). Driving performance under the influence of drugs: Rationale for, and application of, a new test. *British Journal of Pharmacology, 18*, Suppl 1: 121S-129S.

Risser, M.R., Ware, J.C., & Freeman, F.G. (2000). Driving simulation with EEG monitoring in normal and obstructive sleep apnea patients. *Sleep, 23*(3): 393-398.

Schmidt, R.A., & Lee, T.D. (1999). *Motor control and learning a behavioral emphasis (3rd ed.)*. Champaign, IL: Human Kinetics.