Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men

W. Ting, J. Wessel, S. Brintnell, R. Maikala, Y. Bhambhani

Key Words: lifting • physical endurance • work capacity evaluation

Objective. To examine the criterion validity of the Baltimore Therapeutic Equipment (BTE) work simulator by comparing endurance time, oxygen uptake (VO₂), and heart rate measured during real and simulated lifting tasks and to derive a regression equation for predicting actual lifting endurance from measurements on the work simulator.

Method. Twenty healthy male volunteers repetitively lifted and lowered a load of 40 lb using the BTE work simulator and actual weights at a laboratory workstation. Postures, location, and frequency of lifts were kept constant. Endurance (defined as the time taken for the rating of perceived exertion to increase 2 units on the Borg scale) was measured under both conditions. VO₂ and heart rate were also recorded, using standard physiological procedures.

Results. The mean values for endurance time, steady-state VO₂, and heart rate were significantly different between the real and simulated tasks (p < .05). Correlation of endurance time between the two tasks was significant (r = .71, p < .05). Stepwise regression analysis resulted in the following equation for predicting real endurance from simulated time measurements: predicted real time = .34 simulated time + 3.29; r = .71; SE = 1.00 min.

Conclusion. The BTE work simulator tends to overestimate real lifting endurance performance in healthy men. The lower physiological stress during the simulated task suggests a significant difference between the real and simulated loads. Occupational therapists should exercise caution when using the results of the BTE work simulator during functional capacity evaluations.


O ccupational therapists working in occupational rehabilitation programs routinely perform functional capacity evaluations to determine a person's functional performance and employability after injury or illness. The evaluation is used to predict the ability to return to work, to determine direction of the treatment program, to monitor progress, and to provide information for compensation (Hart et al., 1994; Lechner, Roth, & Straaton, 1991; Tramposh, 1992). The use of functional capacity evaluations can prevent a potentially productive person from losing his or her working life prematurely and protects the workers’ compensation system from considerable abuse (Tramposh, 1992). At present, a variety of functional capacity evaluations are available; they vary in the choices of measuring instruments, length of assessment, determination of testing end-points, nature of work activities chosen, and standards of practice (Gibson & Strong, 1997; King, Tuckwell, & Barrett, 1998; Lechner et al.,
Measurement of functional capacity has evolved as a result of great difficulties encountered in translating medical impairment into functional limitation (Abdel-Moty et al., 1993). To ensure proper return-to-work decisions on the basis of functional capacity evaluations, the key issue is the match between the client’s functional capacity and the critical demands of the actual job (Menard & Hoens, 1994). The best method to achieve this match is conducting a situational assessment of how the person functions in an actual work environment. Unfortunately, because of practical constraints, such as safety, limitations of time and personnel, cooperation with employers, and difficulty in quantifying the measurement in a real work situation, situational assessment often is not used. Instead, therapists set up workstations (Smith & Baxter-Petralia, 1992) or use commercially available devices or work samples (Matheson & Niemeyer, 1986) to simulate real work tasks. Clarity, objectivity, and work relevance are critical in the evaluation of functional capacity with simulated work tasks (Isernhagen, 1990). Sometimes, the simulated tasks are performed under optimal measurement conditions, but they may not reflect actual job demands.

The Baltimore Therapeutic Equipment (BTE) work simulator is a commercial device designed for the evaluation of functional capacity (Bhambhani, Esmail, & Brintnell, 1994). This space-efficient equipment allows users to simulate most upper-limb motions and tasks involved in job activities (Curtis & Engalitchef, 1981). Occupational therapists commonly use the work simulator in work evaluation and work-hardening programs. In the past few years, several studies have reported the use of the BTE work simulator in evaluation and treatment of work rehabilitation. These studies have established normative data for physiological responses of three simulated tasks (wheel turn, push–pull, overhead reach) in healthy men and women (Bhambhani et al., 1994; Esmail, Bhambhani, & Brintnell, 1995), grip and wrist flexion strength (Anderson, Chanoski, Devan, McMahon, & Whelan, 1990), and metabolic and hemodynamic responses to various simulated activities in cardiac patients (Wilke, Sheldahl, Dougherty, Levandoski, & Tristani, 1993).

As well, researchers have examined the reliability and validity of the BTE work simulator. Good test–retest reliability for grip strength and wrist flexion has been reported by Anderson et al. (1990) ($r = .91$ and .98) and Trossman, Suleski, and Li (1990) (ICC = .98). Kennedy and Bhambhani’s (1991) study also demonstrated significant correlations between repeated trials at three intensity levels by measuring oxygen uptake (VO$_2$) ($r = .74$–.87) and heart rate ($r = .59$–.78) during real and simulated work in healthy men. However, in a technical study, Coleman et al. (1996) examined the consistency of resistance provided by 12 BTE work simulators. Regardless of weight levels, the simulators did not provide a constant resistance as reflected by the weight-drop time variations. These findings imply that resistance of the simulator can vary within and across the machines. Furthermore, Kennedy and Bhambhani reported that the VO$_2$ and heart rate during a high-intensity simulated task were significantly lower than that of the real task, thereby questioning the criterion validity of this instrument. Wilke et al. (1993) also reported significantly lower metabolic responses in the BTE simulated tasks compared with the real work tasks in cardiac patients. This result leads one to question whether the BTE work simulator can be used to accurately measure endurance performance. If the BTE simulated tasks appear “less strenuous,” measurement of continuous work performance on the simulator may differ from the actual job performance.

The accuracy of evaluating and predicting return to work depends on how well the BTE work simulator can actually simulate the work tasks: The greater the match between the BTE simulated tasks and the real tasks, the more likely the results will apply to a real job situation. The objectives of this study, therefore, were the following:

1. To evaluate the criterion validity of the BTE work simulator in the measurement of endurance by comparing the length of time that healthy men can sustain floor-to-bench lifting under real and simulated conditions at the same intensity
2. To derive a regression equation for predicting actual lifting endurance from measurements on the work simulator.

Mean comparisons for VO$_2$ and heart rate between the real and simulated tasks were also performed to provide further insight on the validity of the BTE work simulator.

**Method**

**Participants**

For this study, written informed consent was obtained from 20 healthy male volunteers 18 years to 37 years of age. Means and standard deviations were 25.4 ± 6.1 years for age, 1.73 ± 0.1 m for height, 69.9 ± 9.8 kg for body mass, and 23.4 ± 3.0 units for body mass index (weight/height$^2$). Persons in this age range were selected because they fall within the group that incurs injuries in the industrial sector (Mital & Pennathur, 1999). Most study participants were university students. Eighty percent were actively involved in various sports and physical activities. The duration of exercise participation ranged from 1 hr to 15 hr per week. All participants were screened for contraindications to exercise with the Physical Activity Readiness Questionnaire (Canadian Society for Exercise Physiology,
Men with known history of medical and orthopedic conditions were excluded. The study procedure was approved by the Human Ethics Review Committee at the University of Alberta.

Procedure

Each participant completed three testing sessions over a 2-week period. Each session lasted about 1 hr and was carried out at a room temperature of 20°C to 23°C. Lifting was studied because it is a common work task and is included in the U.S. Department of Labor’s listing of 20 physical demands of work (Lechner et al., 1991). A load of 40 lb, which is classified as “heavy” according to the Employment and Rehabilitation Institute of California’s version of the Department of Labor’s system (Matheson, Ogden, Violette, & Schultz, 1985), was chosen.

A workstation was set up in the laboratory, using actual weights to perform the real work task. Participants were asked to bilaterally lift and lower a plastic crate (13 in. x 13 in. x 13 in.) weighing 40 lb through a vertical distance of 30 in. between the floor and a bench. For the simulated task, attachment #191 of the BTE work simulator was used. The shaft height and head position of the simulator was controlled such that participants lifted through the same origin and destination as the real task. The resistance was adjusted to a torque of 60 in.-lb to simulate the weight and mechanics and asked to select a work pace at which they felt most comfortable without becoming unusually tired. The pace was recorded as number of repetitions per min and was used in subsequent sessions.

Sessions 2 and 3: Real and simulated work tasks. In these two sessions, participants were asked to perform the real or simulated floor-to-bench lifting and lowering tasks in random order. The self-selected pace determined previously was used and kept constant in both the real and simulated tasks via visual cues on an electronic timer. To quantify endurance, the time during which participants lifted and lowered the 40-lb load repetitively until the rating of perceived exertion (RPE) increased by 2 units was measured. The RPE was recorded at rest and at the end of every minute of lifting with the Borg (1982) scale. The Borg scale is a range of scores from 6 (rest) to 20 (peak effort) that gives a subjective rating of how hard the task is during the test. The scale is carefully constructed, with the following verbal descriptors anchoring the odd numbers: 7 = very light; 9 = very light; 11 = light; 13 = somewhat hard; 15 = hard; 17 = very hard; 19 = very, very hard. A test–retest reliability of .80 during progressive exercise has been reported for the RPE responses (Skinner, Hustler, Bergsteinaova, & Buskirk, 1973). A multiple correlation of .84 was observed between the RPE and heart rate, VO₂, pulmonary ventilation, and blood lactate, demonstrating the validity of the RPE as a measure of exercise intensity (Pollock, Jackson, & Foster, 1986). Typically, it takes a few minutes for the integration of central and peripheral inputs that arise from body movements to obtain an accurate RPE (Birk & Birk, 1987). A steady state is important to achieve in order to reflect the physiological stress experienced by the participant. Hence, the RPE recorded at the end of the 4th min of lifting was used as a reference point, and when the value increased by 2 units from this value, the task was discontinued. A pause time limit between successive lifts (< 30 sec) was also set when the participant felt tired and slowed down his pace.

External motivation was controlled by using the same instructions for both tasks. To ensure accurate determination of RPE, the Borg scale was explained to each participant, using modified instructions from Morgan (1981) and Noble and Robertson (1996). To simulate the real world where workers often take rest breaks on the job, participants were asked to repeat each task using the same procedures with a rest break between the two trials. The second trial was commenced only when the standing heart rate

2Model MET 300, Cybex upper body exerciser, Lumex Inc., 2100 Smithtown Avenue, Ronkonkoma, New York 11779.
3Sport Tester, Model 3000, Polar Key, Professorintie 5, FIN-90440 Kempele, Finland.
4MMC, Horizon; Sensormedics, Inc., 22705 Savi Ranch Parkway Yorba Linda, California 92887.
returned to resting level (± 5 bpm). The rest time between the two trials was recorded for each task.

During the tests, VO₂ and heart rate were monitored according to the procedure described previously. The steady-state values measured during the 4th min were used for analysis. In addition, the VO₂ and heart rate measured for the real and simulated work tasks were expressed as a percentage of their respective peak values obtained during the incremental arm cranking test. The metabolic equivalent (MET) of each task was calculated by dividing the relative VO₂ (mL·kg⁻¹·min⁻¹) by 3.5 to indicate the energy cost of the task as a multiple of the resting energy cost (McArdle et al., 1996). Oxygen pulse (ml/beat) was calculated as the ratio between the absolute VO₂ and heart rate.

**Statistical Analysis**

Two-way analysis of variance (ANOVA) with repeated measures (tasks by trials) was used to examine differences between the means of the two tasks and two trials for each dependent variable. Because of the number of univariate analyses performed, the Bonferroni adjustment was made (Ottenbacher, 1991). To establish validity, Pearson product-moment correlations were used to study the relationship between the real and simulated tasks for selected variables. Stepwise regression analysis (Cohen & Cohen, 1983) was used to predict actual endurance performance from the simulated time measurements. In this analysis, endurance time of the real task was entered as the dependent variable, whereas endurance time of the simulated task, BMI, and resting heart rate were entered as independent variables. BMI and resting heart rate were selected because they can be easily measured by occupational therapists. Results were considered significant at an alpha level of .05. All statistical analyses were performed with the Statistical Package for the Social Sciences, Version 8.0 (SPSS, 1998). Because of technical problems with the metabolic cart, data for two participants were removed from the physiological comparisons (which are presented later in Table 3).

**Results**

**Peak Physiological Responses**

Descriptive statistics for the peak physiological responses are presented in Table 1. During the incremental arm cranking tests, none of the participants attained age-predicted maximal heart rate. However, most reached an RPE of 18 on the Borg scale, and all attained an RER ≥ 1.10. These observations suggested that the participants performed at maximal efforts during the tests.

**Comparison Between Tasks**

Comparisons of endurance time, rest time, and RPE during the real and simulated lifting tasks for the two trials are summarized in Table 2. There were significant differences (p < .05) in endurance performance and RPE between the real and simulated tasks; however, no significant differences were observed between the two trials of the same task. Participants performed the real task for a shorter duration compared with the simulated task. The simulated task was perceived as being light, whereas the real task was perceived as somewhat hard. The rest time between the two trials was also significantly longer for the real task compared with the simulated task.

Results of the two-way ANOVA with repeated measures indicated no significant interaction between tasks and trials for any of the physiological variables examined (p > .05). The lack of interaction implied that for both the real and the simulated tasks, the trend observed in the two trials was the same for each task (Keppel, 1973). Examination of the main effects of tasks indicated significant differences for endurance time, VO₂, heart rate, percentage of peak VO₂, percentage of peak heart rate, and MET (p < .001) but not for the RER and oxygen pulse (p > .05). The mean steady-state physiological responses are presented in Table 3. In general, the mean values were higher during the real task by 6% to 34% compared with the simulated task. Examination of the main effects indicated no significant differences for any of these variables between trials of the real and simulated tasks (p > .05).

**Prediction of Lifting Endurance**

When the endurance time was averaged across the first and second trials, a significant correlation (r = .71, p < .05) was found between the real and simulated tasks. This finding suggested that participants with greater endurance during the real task were more likely to perform the simulated task.

**Table 2**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Real Lifting</th>
<th>Simulated Lifting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1 (M (SD))</td>
<td>Trial 2 (M (SD))</td>
</tr>
<tr>
<td>Endurance time (min)*</td>
<td>6.34 (1.44)</td>
<td>5.85 (1.57)</td>
</tr>
<tr>
<td>RPE*</td>
<td>13.30 (2.68)</td>
<td>13.30 (3.03)</td>
</tr>
<tr>
<td>Rest time (min)*</td>
<td>24.15 (14.06)</td>
<td>19.30 (8.14)</td>
</tr>
</tbody>
</table>

*Note. RPE = rating of perceived exertion.

*Significant difference at the .05 level between real and simulated lifting for each trial. No significant differences were observed between the first and second trials of the same task for endurance time and RPE.
Table 3
Comparisons of Mean Steady-State Physiological Responses During the Real and Simulated Lifting Tasks

<table>
<thead>
<tr>
<th>Variable</th>
<th>Real Lifting</th>
<th>Simulated Lifting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1 SD</td>
<td>Trial 2 SD</td>
</tr>
<tr>
<td>Absolute VO₂ (L·min⁻¹)</td>
<td>1.38 (0.24)</td>
<td>1.40 (0.20)</td>
</tr>
<tr>
<td>Relative VO₂ (ml·kg⁻¹·min⁻¹)</td>
<td>20.0 (3.7)</td>
<td>20.3 (3.4)</td>
</tr>
<tr>
<td>Heart rate (bpm)</td>
<td>138.4 (13.3)</td>
<td>142.6 (13.2)</td>
</tr>
<tr>
<td>VE (L·min⁻¹)</td>
<td>43.3 (8.2)</td>
<td>45.5 (9.3)</td>
</tr>
<tr>
<td>Oxygen pulse</td>
<td>10.1 (1.9)</td>
<td>9.9 (1.6)</td>
</tr>
<tr>
<td>RER</td>
<td>0.97 (0.08)</td>
<td>0.95 (0.12)</td>
</tr>
<tr>
<td>%pHR*</td>
<td>81.0 (8.8)</td>
<td>83.2 (8.9)</td>
</tr>
<tr>
<td>%pVO₂*</td>
<td>91.7 (18.8)</td>
<td>93.7 (20.8)</td>
</tr>
<tr>
<td>MET*</td>
<td>5.72 (1.06)</td>
<td>5.80 (0.97)</td>
</tr>
</tbody>
</table>

Note. VO₂ = oxygen uptake; VE = minute ventilation; oxygen pulse = ratio between VO₂ in L·min⁻¹ multiplied by 1000 and heart rate; RER = respiratory exchange ratio (ratio between carbon dioxide production and oxygen consumption); %pHR = percentage of peak heart rate; %pVO₂ = percentage of peak oxygen uptake; MET = metabolic equivalent.

Discussion

Differences Between Real and Simulated Tasks

Endurance should be considered in functional capacity evaluations if the job demands a high degree of repetitive efforts over a continuous period. The present study demonstrated that when the BTE work simulator was used in the evaluation of endurance performance during floor-to-bench lifting and lowering, there was a significant difference in timed measurements from the real task (see Table 2). Endurance time measured on the BTE work simulator was 35% greater than that of the real task. As well, the steady-state RPE was lower and a shorter recovery time required between the two trials during the simulated task compared with the real task. Given that both tasks were carried out in a similar environment with the same postures, location, and pace of lifts, it is postulated that the difference in endurance performance was due to the inherent inability of the BTE work simulator to duplicate metabolic and cardiovascular demands required by the real lifting task.

The present findings are consistent with those of Kennedy and Bhambhani (1991) and Wilke et al. (1993). During physical work, physiological responses, such as VO₂ and heart rate, increase linearly with the work output (McArdle et al., 1996). Measurement of these changes provides indexes of the level of physiological stress imposed on a person (Garg, Rodgers, & Yates, 1992). The fact that VO₂ was lower by 29% and heart rate was lower by 19% during the simulated task compared with the real task, together with the observed lower RPE, suggests that the theoretical lifting load on the BTE was less than the real lifting load of 40 lb. Occupational therapists, therefore, should use caution when judging lifting performance on the basis of results obtained from the BTE work simulator.

One possible reason for the discrepancy between the real and simulated tasks was the design of the lifting attachment on the BTE work simulator. The lifting and lowering of a 40-lb load was a combination of dynamic and static work. The dynamic and static efforts of the simulated task may have been reduced because the motions were facilitated by a rope-and-pulley attachment. A reduction in the dynamic muscular work would translate into a reduced metabolic and cardiorespiratory stress on the participant and result in the lower values of VO₂ and heart rate. Additionally, the reduced amount of static muscular work in the simulated task may have explained the increased endurance time, lower RPE, and shorter rest period between trials compared with the real task. Static work is more fatiguing than dynamic work (Garg et al., 1992; Grandjean, 1988). During static effort, blood vessels to the muscle tissues are compressed, and blood flow is reduced at intensities exceeding 15% of the maximum voluntary contraction (Garg et al., 1992). Under these conditions, the supply of oxygen and substrates cannot meet the muscular demand, resulting in the accumulation of lactic acid and other metabolites that induce fatigue. Grandjean (1988) reported that under similar conditions, static muscular effort will lead to higher energy consumption, raised heart rate, and longer rest periods compared with dynamic work.

Another possible explanation for the difference between real and simulated tasks is that the participants somehow found alternative strategies to relieve fatiguing muscles and change their lifting mechanics. A few men-
tioned that the simulated task was easier because they could use their knees to help their performance. Wheeler, Graves, Miller, O’Connor, and MacMillan (1994) mentioned that the weak link in the biomechanical system while lifting is usually related to the lumbar extensor musculature. When a heavy load is lifted, the lifting mechanics can change to protect the lumbar musculature (Rodgers, 1992; Wheeler et al., 1994). It was possible, therefore, that the setup of the BTE work simulator allowed participants to develop some ways to minimize muscular efforts.

**Differences Between Trials**

The results of this study indicated no significant differences in endurance performance or physiological responses between the two trials of the real and simulated tasks. The lack of difference could be explained by the sufficient recovery time given between the trials. The physiological stress experienced by the participants was high, as the percentage of peak VO2 and peak heart rate were 90% and 83%, respectively, during the real task and 70% for both variables during the simulated task. It is possible that endurance would have been reduced if participants proceeded to the second trial without adequate recovery. In this study, recovery from the lifting tasks was objectively evaluated by allowing the heart rate to return to within 5 bpm of the standing value before the specific task was initiated. This criterion appeared to be a suitable measure of recovery. These results also suggest that occupational therapists should consider the amount of rest given to a client while conducting functional capacity evaluations. Unfortunately, no data are currently available to justify how much rest should be given between trials to allow a safe evaluation without losing the validity of the test. On the one hand, it is important that the functional capacity evaluation be closely related to the actual job demands, particularly in occupations where worker fatigue is of primary concern (Gibson & Strong, 1997). On the other hand, safety is an issue in such evaluations, and the procedure selected should not place unnecessary stress on clients (Isernhagen, 1990). Sound clinical judgment that relies on the therapist’s knowledge, experience, and observation as well as the client’s feedback are therefore important.

**Prediction of Lifting Endurance**

Although there was a significant difference in endurance time between the real and simulated tasks, a significant correlation was obtained between these two variables. This finding suggests that the endurance time varied consistently between the two lifting tasks. About 50% of variance during the real task was explained by performance during the simulated task. The remaining 50% may be explained by other factors such as age, motivation, change in lifting mechanics, perceived exertion, fear of injury, cardiovascular fitness, muscle strength, and previous training or experience. The most likely reason for the modest correlation was that the performance of the participants during the two tasks was quite homogenous, as evidenced from the scatterplot in Figure 1.

It should be noted that the presence of one extreme data point actually increased the correlation between the two tasks. If this point is removed from the statistical analysis, the correlation between these two variables drops to .35, and the common variance is reduced to 12%. These results suggest that measurement of lifting endurance on the BTE work simulator may not provide an accurate prediction for a real lifting task. Another factor that could have influenced this correlation was the criterion used to define fatigue. It is speculated that if the participants were allowed to continue lifting to the point of very, very hard work on the Borg scale (i.e., a rating of 19), there would have been a greater variability in the performance data, thereby leading to a higher correlation and a better prediction.

Ideally, functional capacity evaluations conducted in the workplace should produce valid information about the worker’s performance. However, this ideal is not often feasible because of both practical and safety concerns (Abdel-Moty et al., 1993; Gibson & Strong, 1997). Further research is needed to develop valid equations that will enable occupational therapists to predict lifting performance via a variety of techniques and pacing strategies. As well, research that increases the ability to predict work performance during an 8-hr workday needs to be conducted.

**Conclusion**

The results of this study indicate that measurement of bilateral lifting endurance on the BTE work simulator differs significantly from actual lifting in a laboratory setting for healthy men. Significant differences in mean values were observed for endurance time, VO2, heart rate, and perceived exertion between the real and simulated tasks. Although a significant correlation was found between the two tasks for the endurance time, the common variance between these two variables was only 50%. Therefore, the equation developed for predicting real endurance from simulated endurance performance should be used with caution. Further research is necessary to validate the BTE work simulator for other occupational tasks in different populations so that it can be used with confidence in functional capacity evaluations.

**Acknowledgment**

This article was based on the first author’s master’s thesis in occupational therapy for the Faculty of Rehabilitation Medicine, University of Alberta.

**References**


