Rating of Perceived Exertion, Heart Rate, and Oxygen Consumption in Adults With Multiple Sclerosis

Leona J. Holland, Marcel Bouffard
University of Alberta

Denise Wagner
University of Calgary

The reliability of oxygen consumption (VO₂), heart rate (HR), and rating of perceived exertion (RPE) at three different workloads was examined during an arm cranking exercise task. Nine persons with multiple sclerosis (MS) and confined to a wheelchair each performed two sessions of discontinuous, submaximal aerobic test on an arm ergometer. Comparisons of the test scores and generalizability theory were used to analyze the data. Both HR and VO₂ were found to be reliable measures under the conditions used in this study. RPE at the same workloads was found to be rather unreliable. Overall, the use of RPE as an indicator of exercise intensity instead of HR appears to be unjustified by the results of this study. Therefore, practitioners who want a quick and efficient method of measuring exercise intensity should use HR instead of RPE for persons with multiple sclerosis.

People with multiple sclerosis (MS) are often told to rest and not overexert themselves because they are prone to extreme fatigue. In turn, the fear of fatigue becomes almost unbearable (Schapiro, 1987). As a result, the role of exercise for persons with MS has become somewhat controversial, partly because the meaning of exercise is misunderstood. To many, exercise can be defined as stressing one’s body to the point of pain or overexertion (e.g., “no pain, no gain”). But in persons with MS it has become quite clear that if they exercise to the point of pain or overexertion, fatigue sets in and weakness increases (Schapiro, 1987).

During the late 1950s, Gunnar Borg created a 15-point psychophysical scale (the Borg Scale of Rating of Perceived Exertion) that represented both objective measurements of physiological responses to exercise and the subjective interpretations of effort by the individual (Borg & Linderholm, 1967). The level of exertion an individual perceives was found in numerous studies to closely correlate with some of the physiological responses, including heart rate (HR) and oxygen consumption (VO₂) (Morgan & Borg, 1982).
Recent research with subjects who have nonprogressive physical disabilities indicates that rating of perceived exertion (RPE) can be used effectively in monitoring and prescribing physical activity. For example, Birk and Birk (1987) found that the same linear relationship exists for the nondisabled population between perceptions of effort and HR in clients with cerebral palsy and spina bifida who used wheelchairs. Birk and Mossing (1988) explored the predictive validity of the Borg scale with HR and ventilation using teenage subjects with cerebral palsy on bicycle ergometers; they concluded that a moderately high correlation could be established between the physiological factors and RPE.

A scale such as the Borg scale may be a useful tool for exercise professionals and fitness enthusiasts who have MS and at the same time who want to ensure that they do not overexert to promote "MS fatigue." Therefore this study was undertaken to determine the reliability of the Borg scale with persons who have MS. Specifically, the investigation examined the following questions:

1. Do people with MS assign the same RPE value to the same three workloads on two separate test occasions during arm cranking?
2. Do people with MS have the same HR at the same three workloads on two separate test sessions when using an arm ergometer?
3. Do people with MS have the same oxygen consumption at the same three workloads during two different test sessions on the arm crank?

Methods

Subjects

Nine adults (3 males, 6 females) with MS, who ambulated primarily with either a manual wheelchair (WC) or an electric scooter or wheelchair, provided written and informed consent to participate in this study. At testing time all the participants had been carrying out a regular exercise program 2 or 3 days a week for at least 6 months. Pertinent characteristics of the subjects are provided in Table 1.

Instrumentation and Measurement

All subjects performed two identical discontinuous tests at the same time of day 5 days apart on a Monark Rehab Trainer. Subjects were seated on a chair fixed to the table on which the arm crank rested. Seat height was adjusted to ensure comfort and efficient cranking, with the handle axis of rotation at shoulder height. Upright posture was maintained with hip straps and back support to ensure that the hips, knees, and ankles were flexed at 90°.

Oxygen consumption was continuously monitored and recorded every 30 seconds during the test sessions using an automated metabolic measurement cart (Sensormedics Model MMC Horizon). The exercise electrocardiogram was monitored with the leads in the CM5 position and was interfaced with the metabolic cart.

The subjects performed a 3-min warm-up, with no resistance, at their desired speed. This was followed by a 5-min rest. Three 3-min work bouts were performed in a random order of intensity, each followed by 5 minutes of recovery. Subjects exercised at easy, moderate, and hard workloads at a target speed of 50 revolutions per minute (rpm). The cadence was guided by the use of a metronome, and rpm was recorded every minute.
Table 1

Characteristics of the Subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sex</th>
<th>Age (yr.)</th>
<th>Weight (kg.)</th>
<th>Yrs since MS diag.</th>
<th>Primary mode of ambulation</th>
<th>Medication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>60</td>
<td>78.3</td>
<td>22</td>
<td>Manual WC*</td>
<td>None</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>33</td>
<td>72.1</td>
<td>15</td>
<td>Scooter</td>
<td>Lioresal^a</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>36</td>
<td>53.6</td>
<td>20</td>
<td>Scooter</td>
<td>None</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>46</td>
<td>64.7</td>
<td>14</td>
<td>Manual WC</td>
<td>Lioresal</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>34</td>
<td>75.4</td>
<td>6</td>
<td>Manual WC</td>
<td>None</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>41</td>
<td>83.6</td>
<td>4</td>
<td>Manual WC</td>
<td>Lioresal prolod^b</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>65</td>
<td>73.7</td>
<td>26</td>
<td>Manual WC</td>
<td>None</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>37</td>
<td>63.7</td>
<td>12</td>
<td>Manual WC</td>
<td>None</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>37</td>
<td>55.7</td>
<td>12</td>
<td>Electric WC</td>
<td>Lioresal</td>
</tr>
</tbody>
</table>

| M      | 43  | 69.0      |
| SD     | 11.6| 10.2      |

^a A muscle relaxant whose adverse effects may include hypotension (<10%); ^b For hypothyroidism therapy which should not affect exercise performance, heart rate, or ventilation. *WC = wheelchair.

The resistance used during the study was adjusted for each subject. All the subjects were regularly exercising under controlled conditions in a physical fitness center for persons with physical disabilities. The information collected by the staff was used to select three workloads (easy, moderate, hard) that were likely to produce HRs of approximately 110, 125, and 140 beats per minute, respectively. From this information, weights of 25 grams to 1.5 kilograms were used to set the appropriate intensities.

Rating of perceived exertion was requested and recorded during the last 15 seconds of each minute of exercise. Subjects were given standardized instruction (modified from Birk & Mossing, 1988) prior to the warm-up on the first test session and then were reminded of the instructions on Day 2 of the investigation. An experimenter obtained the rating of perceived exertion by placing a pointer to the side of the chart and slowly moving it along the length of the scale. Pointer movement stopped at the sound or nod of the subject, and the rating was verified verbally by one of the researchers.

Results

Two complementary viewpoints guided the data analysis strategy. Our first concern was to generate knowledge that characterizes persons with MS in general without reference to any particular person. To find out what happens in general, generalizability theory (Brennan, 1983; Cardinet & Tourneur, 1985), an extension of reliability theory, was used. This type of analysis is representative of the nomothetic approach.

We also recognize that it is not possible to infer relations at the individual
level from relations found at the level of aggregates unless we assume functional homogeneity of subjects (Asendorpf, 1990; Bakan, 1955; Thorngate, 1986). Stated another way, without more information it is difficult to infer consistency at the person level from a generalizability coefficient. However, for practical decisions the most useful source of information is the extent to which a person obtains the same scores under the same conditions over time. For this reason, raw scores for each subject are reported. Further, using this strategy, the functional nonhomogeneity of the subjects could be detected by inspecting the individual change scores (Asendorpf, 1990).

Nomothetic Results

Generalizability theory (Brennan, 1983; Cardinet & Tourneur, 1985; Cronbach, Gleser, Nanda, & Rajaratnam, 1972) was used to determine the extent to which the subjects’ scores obtained in the present study could provide information about other scores obtained under identical workloads at different occasions. Using this theory, it was possible to estimate the relative magnitude of different sources of variance by using analysis of variance techniques. These sources of variance are called facets in generalizability theory and are analogous to factors in experimental designs. Depending on the particular intent of the researcher, these facets can be divided into facets of differentiation and facets of generalization (Cardinet, Tourneur, & Allal, 1976). In our study, people was the facet of differentiation because our intent was to differentiate the relative position of people on a scale (e.g., HR, VO₂, RPE).

On the other hand, the extent to which any observation is subject to certain sources of variations (e.g., occasion, workloads) limits the generalizability of the results. These sources of variation were labeled facets of generalization and are analogous to error variance in classical measurement theory. Further, as our intent was to generalize only to the three selected workloads, this facet was considered fixed for the purpose of analysis.

The GENOVA software package (Brennan, 1983) was used to calculate estimates of variance components for HR, VO₂, and RPE. Due to sampling error, some estimates of variance components were negative, although variances, by definition, must be non-negative (Cronbach et al., 1972). Negative variance components were set equal to zero for calculating the generalizability coefficients (Cronbach et al., 1972). These coefficients were obtained using the following formula (Design 4 from Cardinet et al., 1976, p. 130):

\[ \rho^2 = \frac{\sigma_p^2 + \sigma_{pi}^2 \bar{t}_i}{\sigma_p^2 + \sigma_{pi}^2 \bar{t}_i + \sigma_{po}^2 \bar{t}_o + \sigma_{pio}^2 \bar{t}_i \bar{t}_o} \]

where \( \rho^2 \) = generalizability coefficient; \( \sigma_p^2 \) = variance due to persons; \( \sigma_{pi}^2 \) = variance component due to the interaction of person by intensity; \( \sigma_{po}^2 \) = variance component due to the interaction of person by occasion; \( \sigma_{pio}^2 \) = variance component due to the interaction of person by intensity by occasion (confounded with the error term); \( n_i = \) number of intensities; \( n_o = \) number of occasions.

Data for Subjects 1 and 6 were not included in this analysis (thus \( n=7 \)), because they both had problems maintaining the target speed during the hard intensities of the second test session (severe coughing and loss of upright posture, respectively).
Estimates of variance components are presented in Table 2. For both HR and VO$_2$, the percentage of variance accounted for by person, intensity, and the interaction of person by intensity were large. The same remarks are true for RPE except for the person-variance component, which accounted for 1% of the total variance. This means that if the RPE scores obtained by each person across workloads and occasions are averaged, the variance among people will be very small. Finally, the person-by-occasion-variance components for HR and VO$_2$ were relatively small, while this component accounted for 12% of the variance of RPE.

Table 2

<table>
<thead>
<tr>
<th>Effect</th>
<th>Heart rate</th>
<th>Oxygen consumption</th>
<th>Rating of perceived exertion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VC*</td>
<td>%</td>
<td>VC</td>
</tr>
<tr>
<td>Person (P)</td>
<td>452.540</td>
<td>57</td>
<td>16.447</td>
</tr>
<tr>
<td>Occasion (O)</td>
<td>-116.912</td>
<td>0</td>
<td>-0.134</td>
</tr>
<tr>
<td>Intensity (I)</td>
<td>128.333</td>
<td>16</td>
<td>3.307</td>
</tr>
<tr>
<td>PO</td>
<td>53.749</td>
<td>7</td>
<td>.419</td>
</tr>
<tr>
<td>PI</td>
<td>153.167</td>
<td>19</td>
<td>2.506</td>
</tr>
<tr>
<td>OI</td>
<td>-4.305</td>
<td>0</td>
<td>.016</td>
</tr>
<tr>
<td>POI, e</td>
<td>5.758</td>
<td>1</td>
<td>.069</td>
</tr>
</tbody>
</table>

*VC = variance components. Negative variance components estimates were set equal to zero to calculate the percentage of variance accounted by negative components.

The generalizability coefficients for HR, VO$_2$, and RPE were 0.95, 0.99, and 0.69, respectively. These values indicate that measures differentiating the subjects' status at three workloads, whatever the occasion, were acceptable for both HR and VO$_2$. However, RPE values could not adequately differentiate the subjects' scores. This last finding is not surprising, considering that the person-variance component accounted for 1% of the total RPE variance. Whether this small percentage of variance is due to sampling error remains to be determined.

A different approach was to estimate the precision of an individual measurement or, conversely, the error of measurement (Rowley, 1989). For this purpose, the absolute error variance ($\sigma_{\delta}^2$) was computed by adding the following random effect variance components: $\sigma_o^2 + \sigma_{po}^2 + \sigma_{oi}^2 + \sigma_{pio,e}^2$ (Brennan, 1983, p. 72). The absolute error variance was then used to calculate 95% confidence intervals. These confidence intervals were ±10.16, ±0.91, and ±1.88 for HR, VO$_2$, and RPE, respectively. Compared to the variance of differentiation (the numerator of Equation 1), the 95% confidence interval was large for the RPE. This suggests that for RPE the examined observed scores contain a relatively large amount of error.

It must be underscored that the generalizability coefficients and the absolute error variances are aggregate statistics. They do not indicate whether some individuals are more consistent than others. For this purpose we need to examine the results obtained by each person individually.
Idiographic Results

For each subject, comparisons between the two test sessions for the 3rd minute only at the easy, moderate, and hard intensities were calculated for RPE, HR, and VO₂. The raw scores, as a function of workload and trial, were plotted for each person (see Figures 1 to 3).

Rating of Perceived Exertion. Figure 1 reveals that for Subjects 1, 4, and 5 the difference between RPE values at the same workload were either 0 or 1. These individuals were fairly consistent across sessions. On the other hand, all the other subjects had at least one difference score greater than 1. These preliminary results suggest that some individuals may be much more consistent across sessions. Future work should determine whether there is variability in consistency across individuals and, if so, the correlates of consistency. If some individuals are very consistent, this would support the notion that RPE could be trusted as an indicator of exercise intensity.

Further, Figure 1 shows that the subjects reported the same RPE value at

![Graph showing RPE scores for 9 subjects for two trials at easy, moderate, and hard intensities.]

Figure 1 — Rating of perceived exertion scores for 9 subjects for two trials at easy, moderate, and hard intensities.
the same workload on 15% (4/27) of the trials. An additional 48% (13/27) of the difference scores were within ±1 RPE. Difference scores of 2 or 3 RPE were observed on 37% (10/27) of the trials. Thus it appears that RPE should be employed with caution for persons who have MS, since it varied by two or three units on 37% of the trials completed at the same workload on two occasions.

**Heart Rate.** Figure 2 indicates that Subjects 1 and 7 were more consistent than the other subjects. However, the data from Subject 1 should be interpreted with caution because this subject was coughing during the hard workload of Trial 2. Overall, the variability in response was similar across subjects. HR values obtained during the two test sessions are presented in Figure 2. Eleven of the 27 HR values (41%) varied by ±5 beats/minute or less, while 33% (nine scores) had differences of 6 to 10 beats/minute. The remaining seven HR responses (26%) had differences of more than 10 beats per minute. Thus there seems to be relatively minimal variations in the subjects’ HR scores between sessions.

**Oxygen Consumption.** Figure 3 shows that subjects who had low oxygen consumption (Subjects 3, 6, 7, and 8) also demonstrated relatively small variability
Figure 3 — Oxygen consumption scores for 9 subjects for two trials at easy, moderate, and hard intensities.

in observed scores. Conversely, the high oxygen consumers had relatively more varied scores between two trials. VO₂ measures at each occasion are shown in Figure 3. At each of the three intensities, all the scores varied between ±1.8 ml/kg/min, except for Subject 1 during the hard workload. It was during this high intensity bout that Subject 1 experienced severe coughing. Thus, overall variation in oxygen consumption between sessions appears to be nominal.

**Discussion**

The results of our study clearly indicate that both HR and VO₂ are reliable measures under the conditions used in our study. Further, RPE appears to be an unreliable measure for persons with MS.

The relative unreliability of RPE compared to HR and VO₂ may partly reflect a measurement problem. Conceptually, all these variables are continuous variables. However, the measurement of RPE is rather coarse; it is measured as a discrete variable capable of taking 1 of 15 values. Hence an RPE difference of one unit represents nearly a 7% change between scores. This problem is
relatively nonexistent for HR and VO₂ due to the fact that the measurement procedure is more congruent with the concepts under study. Further, the variance component due to people was very small in our study. Whether this is a result of sampling error remains to be determined by future studies.

One variable that was not monitored in this study and that may have affected the results was core body temperature. Persons with MS cannot tolerate warm temperatures, yet with moderate and high intensity exercise the body temperature rises. It is unlikely that body temperature had a major effect on our results. Rowell (1974) reports that environmental heat and humidity have little effect on performance that is of short duration, perhaps less than 15 minutes. It is only in events lasting more than 15 minutes, or in situations of repeated short work bouts over a prolonged period, that one must be concerned about performance detriments of physiological origin.

In our study we did not limit ourselves to situations in which both room temperature and body temperature are constant. Our intent was to determine the extent to which we can generalize our results in a typical training environment. Thus we wanted to reproduce a situation as representative as possible of a usual training setting for persons with MS. In congruence with generalizability theory, we recognize that different results may have been obtained had we strictly controlled for room and body temperature. Further, all the subjects in this study were regular participants at a training center for persons with physical disabilities, and all were familiar with the Borg scale before the conduction of this study.

Again, different generalizability results could be obtained with other MS subjects who are not regular exercisers or are not familiar with the Borg scale. A major conceptual advance of generalizability theory, contrary to classical true score theory, is that it recognizes that there are multiple sources of error variation and different universes of admissible observations. Generalizability theory does not presume some universally applicable definition of intensity or occasion (Brennan, 1983). Universes of admissible observations could be defined differently by different persons. Hence, different definitions of the universe of admissible observations may lead to different results in a generalizability study. Caution must be exercised in generalizing to other universes.

Four subjects had HRs that were slightly higher (1 to 8 beats/minute) during the medium intensity than during the hard workload in the first test session. Three of these individuals experienced the same response during Trial 2 (Subjects 2, 5, and 6). Although Subject 6 had posture problems during the second session, she experienced no setbacks during Trial 1. All three subjects performed the workouts in a different sequence, thus order is not considered to be a contributing factor. It is not known what caused the nonmonotonic data for nearly half of the subjects. Perhaps changes in the test protocol (e.g., longer work bouts) would reduce or eliminate this nonmonotonic response.

Conclusion

In conclusion, both HR and VO₂ were found to be reliable measures under the conditions used in this study. Further, RPE at the same workloads is less reliable for persons with multiple sclerosis. Overall, the use of RPE as an indicator of exercise intensity instead of HR appears to be unwarranted by the results of this
study. Adults with MS who want a quick and efficient way to self-monitor exercise intensity should prefer HR instead of RPE.

References


Notes

1 The obvious exception to this rule is when the generalizability coefficient is equal to 1. This rarely happens in practice.
2 These variance components were obtained by assuming that all facets were random. For details, see Brennan (1983, p. 70).

Acknowledgments

The authors would like to thank the clients from the Rick Hansen Centre who enthusiastically participated in this study, as well as Ewen R. Nelson for his computer graphics work.