A comparison of the visual analogue scale and modified Borg scale for the measurement of dyspnoea during exercise

RACHEL C. WILSON AND P. W. JONES
Department of Medicine 1, St George's Hospital Medical School, London

(Received 17 March/16 June 1988; accepted 11 July 1988)

SUMMARY
1. The intensity of breathlessness during exercise was measured in ten normal subjects using a visual analogue scale (VAS) and a Borg scale to compare the use of the scales and their repeatability, both within the duration of a period of exercise and between tests. For each scale, subjects performed two exercise tests separated by a period of 2-6 weeks. Each exercise test consisted of two cycles of progressively increasing and decreasing workload.
2. All subjects felt confidently able to use both scales to quantify their feelings of breathlessness exclusively of other sensation. Equal preference was expressed for use of a particular scale.
3. With both scales there was a large intersubject variation in the relationship between dyspnoea score and minute ventilation ($V'_E$) ($P<0.01$), and in the range of the scale used.
4. There was a good correlation between the VAS and Borg scores at each level of $V'_E$ ($r^2 = 0.71$), but the VAS score was used over a wider range than the Borg score.
5. The relationship between $V'_E$ and the dyspnoea score measured by the two techniques was predominantly linear. The mean $r^2$ for VAS score/$V'_E$ was 0.68 ($sd$ 0.19) and for Borg score/$V'_E$ the mean $r^2$ was 0.75 ($sd$ 0.13).
6. The relationships VAS score/$V'_E$ and Borg score/$V'_E$ were unaffected by the direction in which the workload was varied ($P>0.05$).
7. $V'_E$, measured at each work rate, did not differ between the two cycles ($P>0.05$) or between the 2 days ($P>0.05$).
8. With both scales, the slope of the $V'_E$-breathlessness relationship was slightly higher during the second half of the exercise compared with the first ($0.05 < P < 0.01$).
9. The scores with both scales were lower in the second test compared with the first ($P<0.01$): Borg 16% lower, VAS 27% lower.

INTRODUCTION
The sensation of breathlessness is a major link between a patient's lung disease and his resulting disability. There is a need for a reliable measure of dyspnoea that will allow quantification within studies and may allow comparisons between studies. In recent years, two scales have principally been adopted for the subjective measurement of breathlessness. These are the visual analogue scale (VAS), a horizontal line with two anchor points, one at each extreme [1-3], and the Borg scale of perceived exertion [4], which has been modified for breathlessness measurement [5], and has 12 points, 10 of which have verbal descriptors.

The aim of this study was to investigate the manner in which these two scales were used to estimate breathlessness in normal subjects during exercise, and to compare their stability within the period of an exercise test and their reproducibility over time. Normal subjects were used rather than patients, in order to avoid any confounding effects due to possible variations in lung function.

METHODS

Subjects
Ten healthy young volunteers were used (five male and five female) aged between 22 and 33 years. All gave informed consent. Subjects had no history of cardiorespiratory disease and all had normal values of forced expiratory volume in 1.0 s and forced vital capacity. The mean forced expiratory volume in 1.0 s/forced vital capacity ratio was 83% ($sd$ 7.0).
Exercise protocol

Breathlessness was measured during a 32 min exercise period on an electrically braked cycle ergometer (Lannoy). During the first 4 min of exercise the subjects performed unloaded pedalling. The load was then increased by 15 W increments every minute to a maximum of 105 W, then decreased by 15 W steps to 0 W. This cycle of increasing and decreasing workload was immediately repeated.

Objective measurements

Expired gas was collected using a no. 2700 Hans Rudolph valve with mouthpiece and noseclip. Minute ventilation ($V_e$) was measured at 20 s intervals using a Gould 9000IV Pulmonary Exercise Laboratory. This consists of a computer-controlled rolling seal spirometer together with an infra-red $CO_2$ analyser and paramagnetic $O_2$ analyser.

Subjective measurements

Subjects were asked to quantify their feelings of breathlessness. This was defined to them as a feeling of an uncomfortable need to breathe rather than any other sensation associated with exercise, such as fatigue or the awareness that ventilation had increased. At the end of each minute of exercise subjects were asked to estimate their level of breathlessness, if any, using either a VAS or a Borg scale. After each recording, the scale was reset to zero to ensure that every estimation was independent of prior estimations. After the exercise, subjects were asked to comment on the appropriateness and ease of use of the scales for measuring breathlessness.

VAS. The VAS consisted of a horizontal row of light emitting diodes of 100 cm length. It was explained that the left hand end of the scale, labelled 'minimum', corresponded to 'not at all breathless' and the right hand end, labelled 'maximum', corresponded to 'maximum imaginable breathlessness'. The subjects indicated their level of breathlessness by progressively lighting up the display using a hand-operated linear potentiometer. Assuming equal distances to represent equal increments in sensation intensity, subjects were asked to position the light at a point relative to the two extremes of the scale corresponding to their perceived intensity of breathlessness.

Borg scale. The scale used was a modified Borg scale of perceived exertion adapted to be appropriate for measuring dyspnoea [5] (Fig. 1). This consisted of a vertical scale labelled 0–10, with corresponding verbal expressions of progressively increasing perceived sensation intensity. A light could be placed at the appropriate position on the scale, by use of a hand-operated rotary switch.

Experimental design

Each subject carried out four exercise tests on separate occasions (two using the VAS and two using the Borg scale). The two scales were tested within 24–36 h of each other (to allow for recovery from the previous test) and again in a similar manner 2–6 weeks later. The order in which the scales were used was randomized between subjects. Before exercise, subjects were familiarized with the apparatus and the nature of the sensation and measurement methods were described to them. Subjects were naive of the design and objectives of the experiments.

Statistical analysis

The results were analysed, using analyses of variance and co-variance. All comparisons between different exercise periods were performed after removal of the inter-subject variance. This was done using a linear modelling technique (GLIM 3.77, Royal Statistical Society, London) which allowed variables or factors to be progressively added to a statistical model. The subjects' identities were first fitted to the model, then tests were performed on the different experimental states. At the simplest level this is an analysis of variance, but the technique also allows for the existence of individual slopes and intercepts in co-variate analysis. Statistical significance was assessed using an $F$-test and significance accepted at either the 1% or the 5% level (as stated). Results are expressed as mean (sD).

RESULTS

Subjective comments

All subjects felt able to quantify their feelings of breathlessness independently of other sensations associated with exercise, although it was commented that other feelings such as leg muscle fatigue could easily influence estimations. The subjects were evenly divided as to their preferred scale.
Comparison of dyspnoea scores

Reproducibility of ventilatory response to exercise

Reproducibility between days. \( V_E \), measured at each work rate, did not differ significantly between the 2 days (analysis of variance \( P > 0.05 \)).

Reproducibility between cycles. \( V_E \), measured at each work rate, did not differ significantly between the two cycles (analysis of variance \( P > 0.05 \)).

Reproducibility between direction of workload variation. Ventilation during the decreasing limb of the exercise cycle was on average 29.8\% higher than during the period in which the work rate was increased (analysis of variance \( P < 0.01 \)).

Range of scale used

Fig. 2 illustrates the range of the two scales used by each of the subjects. The VAS was used over a wider range than the Borg scale. With the VAS, six subjects reached the upper half of the scale, whereas using the Borg scale only two subjects recorded a maximum score above the half way point. The coefficient of determination \((r^2)\) for the VAS and Borg scores was 0.71.

Relationship between subjective scaling and \( V_E \)

There was a significant correlation \((P < 0.01)\) between VAS score and \( V_E \) (mean \( r^2 = 0.68, \text{sd} 0.19 \)) and between Borg score and \( V_E \) (mean \( r^2 = 0.75, \text{sd} 0.13 \)) in all subjects. In two of the subjects there was a significant positive second-order component in the VAS score/\( V_E \) relationship \((P < 0.05)\), i.e., at higher levels of ventilation, the VAS score increased disproportionately more than \( V_E \). In one of these subjects and also in two others, there was a significant positive second-order component in the Borg score/\( V_E \) relationship \((P < 0.05)\). The variance attributable to the second-order component was very small (2.3\% of the total variance in VAS score and 1.1\% of the total variance in Borg score); therefore, for the remainder of the analysis, the linear co-variance was assumed.

Inter-subject differences

The mean VAS and Borg scores were both significantly different between subjects \((P < 0.01)\). There were also significant differences \((P < 0.01)\) in the slopes of the VAS score/\( V_E \) and the Borg score/\( V_E \) relationships between subjects.

Within- and between-test repeatability of breathlessness scores

Effect of direction of work rate change. Each exercise test consisted of two continuous identical cycles of progressively increasing then decreasing workload, each lasting 16 min. The relationships VAS score/\( V_E \) and Borg score/\( V_E \) were unaffected by the direction in which the workload was varied \((P > 0.05)\).

Repeatability between cycles. After removing inter-subject variance, there were no significant differences \((P > 0.05)\) between the mean VAS scores of the two work cycles or between the mean Borg scores of the two cycles, but the slopes of the VAS score/\( V_E \) and Borg score/\( V_E \) relationships were slightly greater during the second cycle than the first \((0.05 > P > 0.01)\) (Fig. 3).

Repeatability between days. After removal of between-subject variance, mean VAS and mean Borg scores were both lower on day 2 than on day 1 \((P < 0.01)\). On day 2 the mean Borg score was reduced by 16\% compared with day 1, while the mean VAS score was 27\% lower on day 2. The slopes of the breathlessness scores vs \( V_E \) were also compared between the 2 days after removal of the inter-subject variance. On day 2 the slope of VAS score/\( V_E \) was lower than on day 1 \((P < 0.01)\), whereas there was no significant difference between the Borg score/\( V_E \) slopes on the 2 days (Fig. 4).

DISCUSSION

In studies involving scaling of a perceived sensation, there is always the possibility that the subjects may not measure the sensation in question. When measuring breathlessness
during exercise, various other sensations such as leg muscle fatigue or the awareness of increased ventilation may influence the subjects' estimation of the sensation. In the present study, to minimize this problem, subjects were given a clear description of the sensation being studied before each test. After performing the tests, the subjects all reported feeling confident that they were able to exclusively quantify their feelings of breathlessness.

The high level of correlation between VAS and $V_E$ and between Borg and $V_E$ indicates that breathlessness was closely related to the level of ventilation, but we confirmed the established observation that there is a large inter-subject variation [2, 6]. This may be due to the subjects' varying sensitivities to the sensory inputs provoking the sensation of breathlessness, or to the manner in which the subjects use the scales. It is not possible to ascertain from this study what proportion of the inter-subject variation was attributable to each of these factors, but it is worth noting that the individual subjects used the two scaling techniques in a broadly similar manner, despite the inter-subject differences. Absolute values of breathlessness are variable within the population, but the scores obtained in this study are similar to those reported by other workers [2, 3] when related to the prevailing level of ventilation.

There were quantitative differences in the way the two scales were used, however. The Borg scale was used over a rather narrower range than the VAS (Fig. 2). This may have been due to the effect of the verbal descriptors in the Borg scale. These might have imposed a certain threshold of sensation intensity at each level which had to be exceeded before proceeding to the next digit, and may have restricted use of the upper part of the scale by our fit young subjects. The Borg scale has an inbuilt ratio bias with a tendency to restrict scores to the lower half and it should be noted that level 5 on the Borg scale (i.e. half the
Comparison of dyspnoea scores

maximum possible score) is labelled ‘Severe’. The descriptors may also increase preferential use of certain digits. The VAS scale has no constricting descriptors, and this may allow the whole scale to be used more freely. Differences between the use of analogue scales and category scales have been observed in other studies of subjective assessments. Subjects who were asked to indicate their level of discomfort during a helicopter ride, preferentially scored within the central region using a category scale, while a scale without descriptors was used in a more progressive manner [7]. A similar observation was made in the estimation of pain [8]. In this context it should be noted, however, that the correlation between Borg score and $V_E$ was a little better than the correlation between VAS and $V_E$.

Important qualities of a reliable scaling technique are its stability within a test period and its reproducibility between tests. We observed, as have others [3], that in some subjects the dyspnoea–$V_E$ relationship may become curvilinear at higher levels of $V_E$. We found that this occurred with both scales, but in a small minority of subjects. Overall, the first-order regression component was overwhelmingly predominant, and this allowed tests of the reproducibility of the scaling techniques using both the mean score and the slope of the dyspnoea score–$V_E$ relationship. Validation for relating breathlessness to the prevailing level of ventilation is given by Adams et al. [3].

Previous workers have observed that estimation of breathlessness using a VAS was moderately reproducible over the period of 1 week [2, 3], although the scale was not so reproducible over the period of 1 year [3]. In the present study, subjects judged themselves to be less breathless when studied on the second occasion, 2–6 weeks later. This was observed with both the VAS and the Borg scale. These results suggest that the scales are relatively reproducible in the short term, but become less so over longer periods. This variation over time may indicate a change in the subjects’ perception of breathlessness or in the way they use the scales, but this will require further study for clarification. In all cases, the variation in the Borg score was proportionately less than that of the VAS.

There is less information available on the within-test stability of dyspnoea scores, although it has been shown that breathlessness estimation using a VAS increased independently of ventilation during the fourth to the sixth minute of steady-state exercise [9]. We also observed a slightly higher breathlessness score relative to $V_E$ throughout the second half of the test (18 to 32 mins), compared with the first half (4 to 18 mins) both with the VAS and the Borg score. If this variation truly reflected a change in the subjects’ perceived breathlessness, this indicates that either a change had occurred in their physiological state (e.g. muscle fatigue, lactate production, etc.), or in the cerebral processes involved in the perception. It is worth noting that the Borg scale again varied to a lesser extent than the VAS.

We also tested the hypothesis that the severity of breathlessness at a given level of exercise may be influenced by the direction in which the workload is varied (i.e. whether preceded by a higher or lower exercise level), however, we observed no difference in breathlessness estimation between ascending and descending workload, and therefore conclude that each estimation was independent of prior estimations. This is an important observation concerning the consistency of the relationship between dyspnoea score and $V_E$, since $V_E$ was higher at each work rate in the decreasing work rate limb of the test compared with the period of increasing work rate.

Our findings appear to favour the use of the Borg score rather than the VAS, because it correlated with ventilation more closely and provided the more repeatable measure. This is in contrast to what might be predicted if the true relationship between breathlessness and ventilation is a continuum. If this assumption is correct, the VAS should have produced the better correlation, since it potentially provides a large number of scoring levels and should allow measurement of breathlessness with considerable precision. In contrast, the Borg score provides a relatively coarse measure since it has a maximum of 12 levels of intensity and on average our subjects only used six. The use of an ordinal or category scale such as the Borg to describe a continuous variable such as breathlessness will lead to measurement errors because of the approximations required to force the magnitude of the perceived sensation into pre-ordained limits or categories. Any errors in the measurement of a variable will lead to a weakening of the calculated correlation between it and related variables. Another factor which may have influenced the strength of the correlations between $V_E$ and breathlessness score measured using the two scales was the scoring range used by the subjects for each scale. The Borg scale was used over a narrower range than the VAS. This should have improved the correlations in the VAS score/$V_E$ relationship compared with Borg score/$V_E$, since a wide range of values favours a higher correlation coefficient. It should be noted that the difference in the correlations between Borg score/$V_E$ and VAS score/$V_E$ was not due to a difference in the number of $x$–$x$ pairs in the two correlations, because these were identical. We conclude that it is difficult to identify a statistical property of the Borg score/$V_E$ relationship that accounts for its better correlation compared with VAS score/$V_E$ and suggest that it may be related to the psychophysical properties of the Borg scale, possibly the use of verbal descriptors.

The limited scoring range of the Borg scale may have contributed to the higher repeatability of breathlessness measured using this scale; however, it should be noted that the repeatability comparisons were not performed using the breathlessness score at any common fixed point, such as a given work rate, but on the slopes of the relationships between breathlessness and $V_E$, since it was judged that $V_E$ will determine breathlessness more closely than work rate. It therefore follows that the factors which allowed the Borg scale to produce a well-correlated description of the breathlessness–$V_E$ relationship may also have facilitated the reproducibility of the relationship when measured with this scale. It is largely beyond the scope of this study to discuss the relative sensitivities of these two scales when used to
measure experimental or disease-induced changes. In principle, the VAS offers a greater potential for identifying very small changes, but this may depend on the study design and analysis. In many studies it will not be possible to control the breathing level, so both of the variables in the breathlessness–\(V_E\) relationship will be changing. When this occurs, the measured outcome variable is likely to be the slope of the breathlessness–\(V_E\) relationship or a derivative of it. Under these circumstances, it is possible that small changes in this relationship may be detected more readily using the Borg scale. This is because the likelihood of a ‘type II’ error (failure to detect a statistical difference because of ‘noise’ in the measurement) may be reduced, since the Borg scale appears to describe the relationship between breathing and perceived breathlessness with a slightly better level of correlation than the VAS.

In conclusion, the two scales were used in a similar manner in relation to the level of breathing and both operated well in this study. From a statistical viewpoint, the VAS has the potential to provide a more reliable and sensitive measure of dyspnoea, since it allows greater precision in the measurements than the Borg scale. Our results, however, showed the Borg scale to be generally more reproducible than the VAS, both within the period of a single exercise test and between tests. The Borg scale also correlated with the level of ventilation a little better than the VAS. These observations lead us to marginally favour the Borg scale for studies on dyspnoea, although the final choice of scale may depend on the purpose of the study.

ACKNOWLEDGMENT
R.C.W. was supported by a grant from the Wellcome Trust.

REFERENCES