Upsala J Med Sci 92: 1–17, 1987

A Psychophysical Study of Work-Related Stress Using Observer Ratings

M. Wangenheim,[1] G. Borg,[2] P. Holzmann[1]

[1]*Bygghälsan Research Foundation, Danderyd, Department of Rehabilitation Medicine, University Hospital, Uppsala, and* [2]*Department of Psychology, University of Stockholm, Stockholm, Sweden*

ABSTRACT

In two experimental series a panel of observers rated the perceived exertion of a worker performing different tasks. The first experiment involved the lifting of boxes of different weights. The second experiment comprised combined static and dynamic work of the assembly-line type during a full working day.

The results show that the observers were able to discriminate between the different loads in all stress situations. Since the loads ranged from low to maximal stress, all were able to make assessments over the full range. Furthermore, the relationship between observer ratings and self-ratings approximated a logarithmic function with the observers tending to overestimate the low stresses while there was closer agreement in the rating of high stresses.

INTRODUCTION

In principle, several different methods and types of measurements can be used to determine physical stress. It is possible to perform "absolute" physical measurements showing what physical forces are involved in the work, the amount of torque, energy consumption etc. Such physical measurements are sometimes feasible, but the calculations can prove to be difficult to perform. Another type of measure consists in physical measurements of the worker's performance (13). In this context use is often made of maximal performances and fatigue curves, or occasionally of the worker's performance at some "suitable" or "preferred" intensity of work.

Physiological measurements of stress in industry are perhaps some of the most common objective measurements. Methods for determining heart rate, oxygen consumption, EMG (with the exception of surface EMG) and the like are generally very reliable (1). Two disadvantages of these methods are, however, that they require apparatus and time, and that the measurements often involve some infringement upon the ongoing working procedure. Thus a work performance can be analysed from the point of view of the amount of energy used on the basis of oxygen consumption or in terms of physical stress on joints, e.g. the elbow or knee joint, or on the back, e.g. intraspinal pressure etc. It may be difficult to see how the different physiological measures of stress should be weighed in relation to each other or what importance should be attached to a specific measurement. An important question is just how representative the measurements are of the usually quite complex procedure being performed. An entirely different type of measurement of physical stress consists in perceptual estimates of the magnitude of strain. In a number of studies made more recently, this type of measurement has proved to be an excellent complement to the other ones (2, 4, 10).

One might think that the effect of physical stress must be measured in physiological variables. This is not correct, however, although it is a rather common notion. By analogy with this line of reasoning, it would only be permissible to use psychological variables for mental or psychological stress. However, as we know, it is common, and often also appropriate, to use physiological units of measurement for mental stress. Similarly, it is often expedient to use psychological variables, e.g. ratings of perceived intensities, as measures of physical or physiological stress.

Some of the disadvantages of using measurements based on mental perception as indicators of work-related physical stress are the fact that some individuals find it difficult to make reliable estimates and that certain types of stress do not manifest themselves in perceptually relevant changes. There are some advantages, however, one being that the individual spontaneously integrates the different stress factors to which he is exposed. Where the physiological approach gives us a number of disparate measurements, here we obtain an integrated perception or experience of the stress, a "gestalt" of the complete stress pattern. If any particular stress is especially critical, it will be emphasized the most. A further advantage in this respect is that the subjectively perceived stress will directly reflect the capacity of the subject and thereby constitute a relative measure.

The methods generally used in perception psychology or psychophysics are the so-called ratio-estimating methods, i.e. methods employing scales with a zero point and equal distances between points on the scale (15, 16). However, in many field studies use is made of simple scales for estimates in the form of category scales, which only allow a ranking of intensities and not "measurements" approaching ratio scale determinations. A review of different methods used in practical situations has been presented by Pearson (12).

In order to reduce the fundamental disadvantages that the conventional rating scales are fraught with, Borg has elaborated a new methodology that produces ratings of a metric order approaching that of physiological measurements. By inserting verbal descriptions of the type "very slight" and "strong" on a ratio scale from 1 to 10 according to the perceptual intensity as determined by special psychophysical and physiological measurements, a measuring scale has been produced which allows quantitative determinations and mathematical computations (5).

The method can be used for ratings by the worker himself when he perceives a physical stress (9, 7), as well as by other individuals observing subjects at work. The fact that ratings by observers based on visual information obtained by watching a subject doing work can produce reliable measurements of physical stress intensities, has been demonstrated by Runesson and Frykholm (14).

The aim of the experiments described below is to determine the relationship between observer rating and self-ratings (by the worker himself) of work-related stresses. The basic idea of the experiments was to expose a worker to a number of different stress-producing situations under well controlled laboratory conditions. These situations included lifting of boxes as well as situations involving a combination of static and dynamic work over an 8-hour period. In each situation both the worker and a panel of observers rated the stress on Borg's scale. All ratings were entirely individual. This was followed by a comparison of the worker's own ratings of the perceived exertion with those of the observers. The evaluations were based on the assumption that the self-ratings are the objectively correct ones since the relevance of rated perceived exertion (RPE) on Borg's scale to physiological stress parameters in connection with the worker's own ratings has already been demonstrated (5).

METHOD

Lifting of boxes

A panel of 6-8 individuals rated the perceived exertion of a worker lifting a box whose weight was unknown to him as well as to the observers. All participating individuals had previous experience with Borg's scale.

The weight of the boxes varied between 1.5 and 22.5 kg in 6-8 progressive increments. At first the box was lifted, using a two-hand grip, only a few centimetres from the floor, with the knees straight and the back bent at an angle of 90^{0}. Next the worker lifted the box to a posture with his back bent forward at a 45^{0} angle. Later on the worker was allowed to lift the box in any manner desired until an upright posture was reached with his arms extended straight forward while holding the box. Finally, the box was again lifted to a 45^{0} forward inclination of the back, at which time the weight range as well as the number of repetitions of each lift was increased to the maximum (8). All lifts lasted 15 seconds from the starting signal until the signal to put the box down was given. Between the lifts the worker rested for 45-60 seconds.

Combined static and dynamic work over 8 hours

The second experiment involved a form of physical stress similar to an actual industrial situation - work at a conveyor belt. The experiment comprised a full working day during the course of which the same load recurred several times in order to elucidate the effect of continual fatigue.

The device used consisted of a conveyor belt that feeds out blocks at optional intervals of time. These blocks, whose outer dimensions are the same for all weights, were moved about 50 cm downwards, whereupon the worker lifted them up again.

The second experiment comprised 4 series lasting 8 hours. Each such series comprised 9 different load situations defined by 3 different weights of the blocks: 0.1, 2 and 5 kg; and 3 different rates of work: 12.5, 25 and 37 blocks per minute. These 9 work loads were varied within each series.

During all experiments the weights were varied quasi-randomly. This means that the increments from lower to higher weights, and vice versa, were not allowed to form progressive series. Nor were the heavy or light weights allowed to accumulate at the beginning or end of the experiments.

RESULTS

Lifting of boxes

The first experimental series produced results that indicated a clear correlation that makes it possible to use Borg's scale in connection with observer ratings, provided that certain points are observed.

The results show that the order in which the weights are presented sometimes influences the rating of the intensity of the effort. The later in the experiment a weight is presented, the heavier it is rated. This may be the result of increasing fatigue as the experiments progressed. Consequently, the recovery period of 45-60 sec probably was not long enough. On the other hand, the panelists' ratings seemed to reflect the stress resulting from the work and were not merely an attempt to estimate the weight of the box. The panel tended to overestimate low loads and underestimated high ones (fig 1). This can be interpreted as a certain degree of overestimation of the worker's movements and underestimation of the weight of the load.

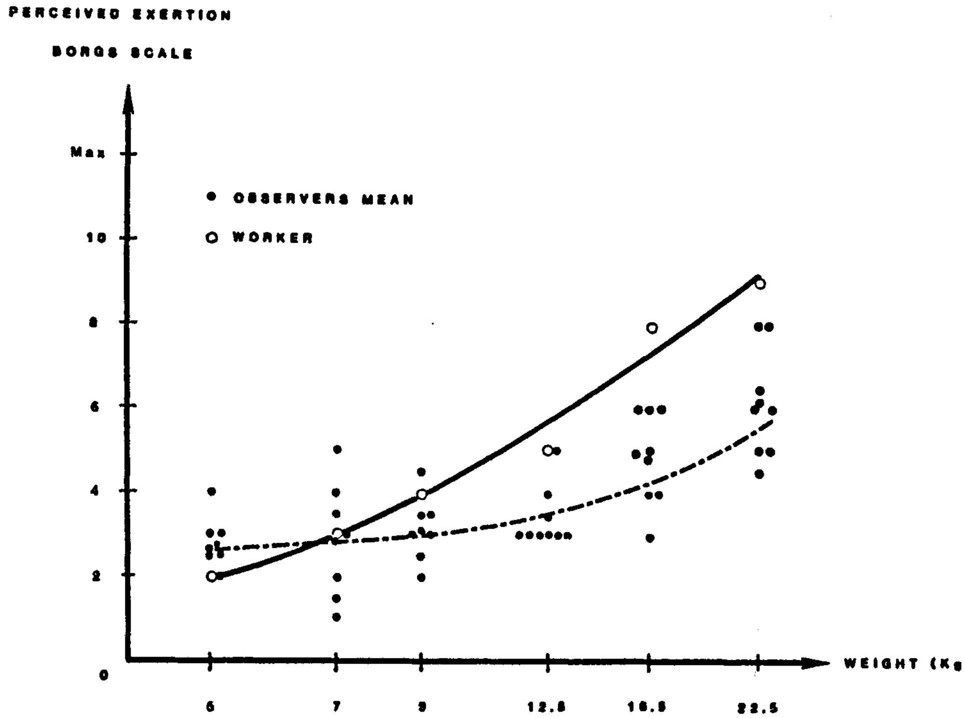

Fig. 1. Correlation between ratings by the observers and by the worker for the lifting of boxes of six different weights in the first three tests in the first experiment.

All of the experiments produced relatively good correlations between the ratings arrived at by the panel and by the worker himself. The regression line of best fit followed a logarithmic function (fig.2), which can be characterized by a formula of the type:

Worker's rating = a x ln (observer rating) + b

a, b = constants

$Y = 6.5881 \ln X - 3.511$

$R = 0.9765$

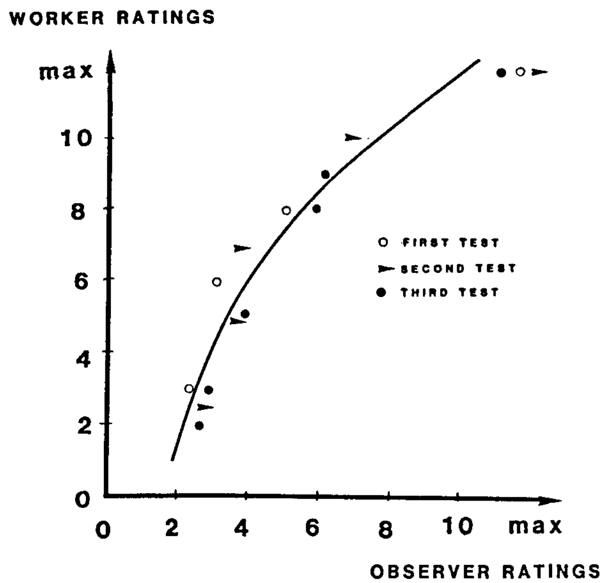

Fig. 2. Correlation between ratings by the observers and the worker himself, on Borg's scale. Summary of the first three tests in the first experiment, lifting of boxes of different unknown weights.

The relationship between the worker's rating and those of the panel in the fourth test is presented in figure 3. A line that best fits the scatter of dots runs a course that is practically parallel to the one obtained for the static work in the first three tests.

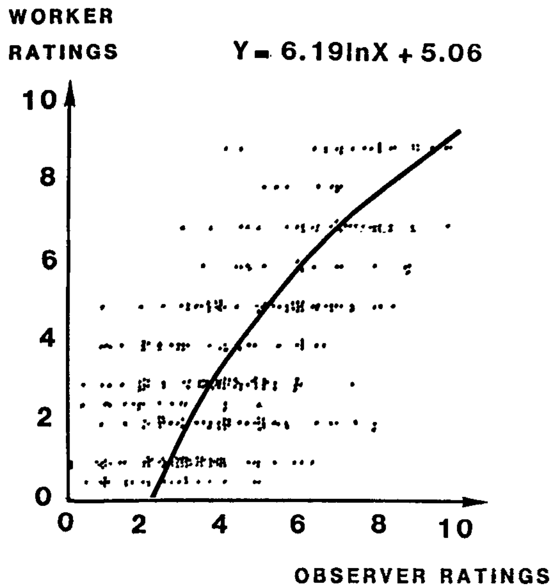

Fig. 3. Correlation between ratings by the observers and by the worker for the lifting of boxes in the fourth test in the first experiment.

If we take a look at the mean ratings of the individual members of the panel, we find considerable variation (table 1). The mean for all members of the panel is 4.34, ranged between 3.01 and 7.58. However, the majority of the panel members used 8-9 points of the 10-point scale. It can be interpreted to mean that the different observers tend to concentrate their ratings within a certain portion of the scale, the differences being influenced by individual factors.

Table 1. Means and dispersions of the ratings by 8 observers and by the worker himself for the 64 load situations (8 weights, 8 repetitions) in the fourth test in the first experiment.

WEIGHT (KG)	OBSERVER PANELISTS' MEAN RATINGS PERS 1	PERS 2	PERS 3	PERS 4	PERS 5	PERS 6	PERS 7	PERS 8	PANEL'S S.D	PANEL'S MEAN RATING	WORKER'S MEAN RATING	WORKER'S S.D
1.5	2.63	2.63	3.38	5.13	1.75	2.69	3.25	2.63	0.92	3.01	0.81	0.35
2.1	2.25	3.38	4.00	5.25	2.50	3.00	2.88	2.63	0.92	3.24	1.44	0.46
2.9	3.38	4.00	5.50	5.63	4.25	3.88	4.00	3.88	0.76	4.32	1.88	0.54
4.0	3.25	4.13	5.75	6.00	3.13	3.63	3.13	1.75	1.33	3.85	2.75	0.43
5.8	3.43	4.71	5.14	5.00	3.43	3.57	3.29	4.00	0.71	4.07	3.29	0.70
8.2	4.38	5.50	7.50	6.25	4.50	4.13	4.50	3.38	1.24	5.02	4.44	0.58
11.3	4.00	5.00	7.75	7.63	5.25	5.56	5.75	6.88	1.24	5.98	5.63	1.22
15.8	5.50	7.50	8.25	9.00	7.88	6.38	8.13	8.00	1.05	7.58	8.00	1.12

A study of the manner in which the worker and the observers rated the same load when it recurred during the course of the experiment shows that the worker is able to discriminate between the different loads considerably better than the panel. Above all, it was difficult for the panel to differentiate between low loads. At high loads the panelist's ratings correlated considerably better with the actual weight of the box and the worker's self-rating.

Combined static and dynamic work over a period of 8 hours

The experiments showed that all subjects were able to clearly discriminate between the different weights. A positively increasing function for the different loads was noted, as well as a fatiguing effect with the passage of time.

On studying the trend in the assessment of weight - rate variants on Borg's scale it was found that, regardless of the rate, the assessment increases with increasing weight. This is well in line with studies made on Borg's scale in connection with its construction. All subjects were able to discriminate between the different weights.

On studying the relationship between ratings on Borg's scale and the actual amount of work done it was found that both the panel and the worker showed a positively rising trend in ratings with increasing work.

A study of the trend with regard to how the same loads were rated when they occurred at different times of the day, shows that the worker rated the low loads more consistently than the panel regardless of the time of day at which they occurred (fig 4). The high loads show a slight tendency to be perceived to be heavier later during the day. The mean ratings increased from 2.1 in the first series to 2.5 in the last one. In contrast, the panel shows a progressive increase in the estimated values at the rates of 12.5 and 37 blocks/min. Here the average value per series rose from 1.4 for the first series to 2.9 for the last one, i.e. almost double the original figure. This indicates a clear tendency to fatigue during the day.

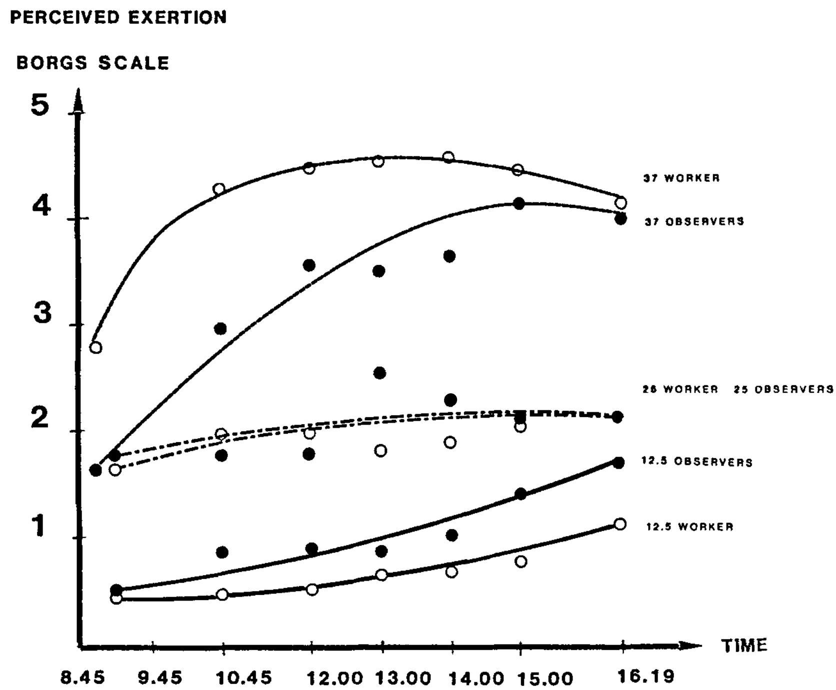

Fig. 4. Correlation between ratings by the observers and by the worker for the different loads over the whole period of work in the second experiment. The assembly-line type work involved lifting of blocks of different weights fed at three rates: 12.5, 25 and 37 blocks per minut.

The ratings show that the work-related stress does not cover as wide a range of variation as was assumed when the experiments were being planned. This is probably due to the fact that the weight of the arms themselves constitutes a large portion of the total load. The worker's self-ratings range between 0 and 6, which means that he also used the same portion of Borg's scale as the observers.

A look at the individual variation in the observers' ratings shows that the average dipersion for all 36 situations was

S.D. = 0.85 (table 2). For the individual situations, S.D. ranges between 0.1 and 1.3. Further, the value of the observer's average ratings ranged between 0.7 and 4.1 on Borg's scale. Thus only about 40 % of the potential of the 10-point scale has been utilized. Individually, the panelists used values between 0 and 7. The design of the experiment was based on a real industrial situation, work at a conveyor belt. Such work is characterized by moderate physical stress. Furthermore, the worker was in very good physical condition. As a result, the work load was concentrated on one portion of the 10-point scale.

Table 2. Means and dispersions of the ratings by 6 observers and by the worker himself for the 36 load situations (9 weight/rate combinations, 4 repetitions) in work involving lifting blocks at a conveyor belt.

COMBINATION: WEIGHT (KG) AND FREQUENCY RATE (BOXES / MINUTE)	MEAN RATING PANEL $\bar{X}$	PANEL'S MEAN DIS-TRIBUTION	MEAN RATING WORKER
0/12.5	0.71	0.63	0.0
0/25	1.03	0.60	0.5
0/37	1.52	0.83	1.5
2/12.5	1.69	0.77	1.0
2/25	1.73	0.72	2.0
2/37	3.04	1.09	3.0
5/12.5	3.46	1.07	4.25
5/25	2.44	0.69	3.25
5/37	4.08	1.33	5.75

Since the complete range of Borg's scale was not used in the execution of the second experiment owing to the fact that very heavy loads did not occur, the results show an accumulation of values on the lower portion of the scale (fig. 5). This, in turn, means that the calculated regression line must be interpreted cautiously, and extrapolating to higher values on the 10-point scale is inadmissible.

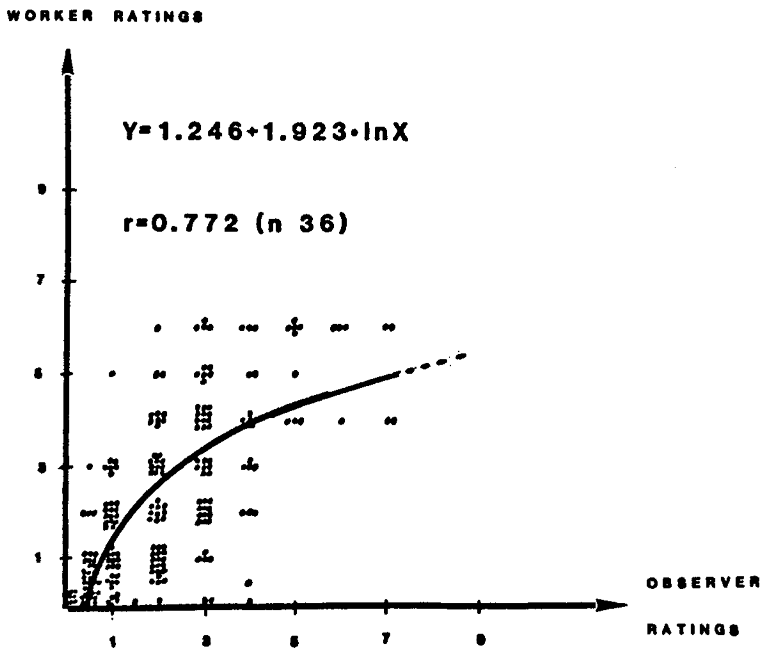

Fig. 5. Correlation between ratings by the observers and by the worker himself on Borg's scale. Summary of the entire second experiment work on a conveyor belt.

DISCUSSION

In all work situations the observers were able to discriminate between the different loads. In cases where the load encompassed the whole spectrum of weights, i.e. weights ranging from very low to maximal, the observers made use of perceptual ratings over the entire scale. This was advantageous to establish a reliable relationship between observer ratings and self-ratings. In other cases, such as work at the conveyor belt, the curve encompassed only a portion of the scale range. Throughout the experiments the curves followed a positively accelerating power function with increasing work loads, as had been demonstrated in Borg's earlier studies (4).

The correlation between the observers' rating and the worker's ratings was not linear, but followed a curved line that could be best described by a logarithmic function. Another finding was that the observers tended to overrate the low loads, whereas the ratings for the higher loads showed better agreement. In principle, with the aid the logarithmic function, it should be possible to figure out how a worker would indicate his own perception of the stress if the observers' rating of the same work is known.

The experiments also showed that the observers tended to rate according to a certain individually coloured pattern. The most noticeable feature was that certain individuals tended to deviate consistently either towards lower or towards higher ratings, so-called "low-raters" and "high-raters". This is a fairly well-known effect mentioned in the literature as the range effect. It can be corrected, if so desired, by calculating the individual deviation from a mean for a number of observers and then correcting all of the individuals ratings by a simple factor so that his mean agrees with the common mean. Another type of personal deviation seemed to be the range used by the observer, i.e. the difference between the highest and the lowest rating indicated. There were individuals who used the complete range of the 10-point scale and others who tended to restrict themselves to a limited area of the scale. Although the observers were tranined in RPE, these differences in the use of the scale could be confirmed, in conformity with the findings of Das (8) and Sury (17).

If a correction is to be made, the best solution would probably be to introduce a test of the raters' ability to use differentiated scales of the Borg type, like the ones that have already been used by him.

A special case presents itself when the "maximum load" rating is indicated on Borg's scale. This increment on the scale is the only one without a numerical value for the magnitude of the load. The purpose of this is to make it possible for the observer to indicate exceptional loads whose occurrence is extremely rare. It is also conceivable to designate loads beyond the capability of the worker as "maximal".

In the light of the experimental results, the suitability of Borg's category scale with ratio properties for use in work analyses based on observers' ratings can be assumed. However, an improvement in this respect could be introduction of a correction factor for the observers' ratings. This factor could be suitably formulated as the common regression line with logarithmic characteristics.

ACKNOWLEDGEMENTS

This study is based on data collected in 1982 within the framework of the ARBAN project, supported by research grants 79/82 from the Swedish Work Environment Fund to Bygghälsan Research Foundation.

REFERENCES

1. Basmajian, I.V.: Muscles alive, chapter 2. William & Wilkins company, Baltimore, USA, 4th edition, 1979.
2. Borg, G.A.V.: Physical performance and perceived exertion. In: Studia psychologia et paedagogica, series altera, Investigations XI, Lund, 1962.
3. Borg, G.A.V & Noble, B.: Perceived exertion. In: Exercise and Sport Sciences Reviews (vol. 11). (ed. J.H. Wilmore). New York. Academic Press Inc., 1974.
4. Borg, G.A.W.: Physical work and effort. Wenner-Gren Center International Symposium Series, volume 28. Pergamon Press, 1977.
5. Borg, G.A.V.: A category scale with ratio properties for intermodal and interindividual comparisons. In: Psychophysical Judgment and the Process of Perception. (ed. G.-G. Geissler & P. Petzold). Berlin: VEB Deutscher Verlag der Wissenschaften, 1982.
6. Borg, G., Herbert, A. & Ceci, R.: Some characteristics of a simple run test and its correlation with a bicycle ergometer test of physical working capacity. Reports from the Department of Psychology, The University of Stockholm, 1984.
7. Corlett, E.N. & Bishop, R.P.: A Technique for Assessing Postural Discomfort. Ergonomics 19, 2: 175-182, 1976.
8. Das, B.: A statistical investigation of the effect of pace and operation on performance rating. Int. J. Prod. Res. Vol 3, No 1: 65-72, 1965.
9. Kirk, N.S. & Sadoyama, T.: A relationship between endurance and discomfort in static work. Unpublished M. Sc. Report, Department of Human Sciences, University of Technology, Loughborough, 1973.
10. Marks, L.E.: Individual differences in perceived exertion assessed by two new methods. Perception & Psychophysics 34, 3:280-288, 1983.

11. Mihevic, P.M.: Sensory cues for perceived exertion: a review. Medicine and Science in Sports and Exercise. 13, 3:150-163, 1981.
12. Pearson, R.G.: Subjective aspects of performance. In: Handbook of Industrial Engineering. (Ed. G. Salvendy). Wiley & Sons, 1982.
13. Rohmert, W. & Laurig, W.: Evaluation of Work Requiring Physical Effort. Institute of Industrial Science, Darmstadt Polytechnic, 1975.
14. Runesson, S.& Frykholm, G.: Visual Perception of Lifted Weight. Report No. 268, Dept. of Psychology, University of Uppsala, 1979.
15. Stevens, S.S.: On the Psychophysical law. Phychological Review 54, 153-181, 1957.
16. Stevens, S.S.: Issues in psychophysical measurement. Phychological Review 78: 426-450, 1971.
17. Sury, R.J.: A comparative study of performance rating systems. Int. J. Prod. Res. Vol. 1, No. 2: 23-38, 1962.

Address for reprints:
Michael Wangenheim
Research Foundation for Occupational Safety and Health in the Swedish Construction Industry
Box 94
S-182 11 Danderyd
Sweden