Health & Fitness Journal
of Canada

Volume 6 December 30, 2013 Number 4

OUR PERSPECTIVE

A brief critique of physiological employment standards for physically demanding public safety occupations.

Roy J. Shephard[1]

Abstract

Background: The time to complete occupational circuits, estimates of maximal aerobic power and simple measures of muscle strength are currently used as employment standards where the safety of the public is judged at risk. *Purpose*: To make a brief critique of the procedures presently used, with a view to identifying on-going weaknesses in assessments. *Methods*: Analysis has focused upon issues of task reactivity, task duration, test components and test accuracy, with particular reference to industrial assessments of maximal aerobic power. *Results*: Test reactivity may exaggerate apparent job requirements. Current tests are of a few minutes duration; extrapolation to longer periods of activity is limited by inter-individual differences, including a possible greater tolerance of prolonged activity in women. Physical effort may contribute less to successful task completion than the mode of approach, which is rarely assessed. The accuracy of physiological measurements of the individual remains a major concern, with errors in the prediction process, effects related to body mass, and substantial intra-individual test-retest variation. *Conclusions*: The concept of objective employment standards for physically demanding public safety occupations is now well accepted. However, practical implementation of individual testing at a level of accuracy to avoid substantial misclassification remains an on-going challenge. **Health & Fitness Journal of Canada 2013;6(4):108-115.**

Keywords: Body mass; Energy costs; Fatigue; Intra-individual variation; Misclassification; Predicted aerobic fitness; Prolonged activity; Task reactivity.

From the [1]Faculty of Kinesiology & Physical Education, University of Toronto, Ontario, Canada. Email: royjshep@shaw.ca

Introduction

In an elegant analysis of physiologically employment standards for a group of physically demanding public safety occupations that included structural and wild-land firefighters, police, correctional officers, and nuclear power emergency personnel, (Jamnik et al., 2013) focused upon task simulation circuits or hybrid evaluations that combined a task simulation circuit with minimal standards of aerobic power; they argued that together these form the basis of a defensible *bona fide* requirement for such occupations. Test batteries of this type have seemingly proven helpful in resolving Human Rights issues concerning the employment of women, older workers, and minority groups where public safety was a concern, and the process has been accepted by the courts, employers and employees as relevant to the experience of front-line work. However, potentially unresolved issues in the practical application of this approach include test reactivity, task duration, unmeasured areas of competence and above all the accuracy of test results.

Test reactivity

Reactivity is well recognized as a complication when measuring habitual physical activity by means of an instrument such as a pedometer (Clemes et al., 2008). The wearing of such an instrument can boost the physical activity of test subjects by 10-20% for a week or more. Although less well documented, a similar phenomenon seems likely when assessing the physical demands of occupational tasks by heart rate monitor or oxygen consumption measuring device. In particular, the rate of performance may be boosted beyond the daily norm in order to impress an observer or to support contract demands.

One safeguard against this danger has been a subjective assessment of the content validity of measurements, with experienced observers indicating whether they agreed that a test provided an appropriate evaluation of the ability to perform a critical physically demanding task. However, a stronger safeguard would be to obtain video records allowing a comparison between the working pace during a test evaluation and that observed during normal operations. If there is indeed a reactive response to testing, this could well be setting minimum performance standards 10-20% higher (or lower) than those needed during actual operations.

Task duration

Many physical ability tests have been of short duration. Times for the completion of job-simulating obstacle courses have ranged from 116 to 255 sec (Farenholtz and Rhodes, 1990 ; Gledhill and Shaw, 1996 ; Lagassé et al., 1985; Rhodes and Farenholtz, 1992). Equally, measurements of oxygen intake have been of relatively short duration, and many tests of muscular strength have required a single brief effort, for example dragging a 68 kg mannequin once over a distance of 15 m. However, there is good physiological evidence that aerobic performance declines exponentially with task duration, and that maximal muscle force declines with repeated efforts. The average person can develop close to100% of their maximal aerobic power for a few minutes, but (depending on muscular strength, body posture and any added thermal stress), this ceiling diminishes with time, and is unlikely to exceed 30-40% of maximal oxygen intake over a working day (Åstrand, 1967 ; Bink, 1962 ; Bonjer, 1968 ; Louhevaara et al., 1986) suggested that in the average worker, the time-related degradation of aerobic power was relatively steep: 63% of the initial value could be sustained for 1 hour, 53% over 2 hours, 47% over 4 hours, and 33% over 8 hours.

Importantly from the viewpoint of personnel assessment, the ability to sustain an oxygen intake close to the maximal value varies substantially between individuals (Tschakovsky and Hughson, 1999). A general equation proposed by (Louhevaara et al., 1986) accounted for only 41% of the inter-individual variation in work times at any given percentage of an individual's maximal oxygen intake. One of the characteristics of the successful marathon runner is an ability to exercise at close to the maximal oxygen intake for 2 hours or more (Costill, 1972) moreover, there is some evidence that women have a greater ability than men in terms of sustaining an ultra-endurance effort (Speechly et al., 1996). Local muscular strength, and thus the rate of accumulation of lactate may be another factor leading to intra-individual differences in tolerance of prolonged effort.

It could be argued that for many safety-related occupations, emergencies are of short duration, and that differences in an individual's ability to sustain a high fraction of his or her maximal oxygen intake are therefore not important. Thus, Tipton et al. (2013) attempted to draw a distinction between brief tasks that were critical in an emergency (but performed rarely) and less demanding routine tasks that nevertheless had to be performed without undue stress to the individual and his or her co-workers. Likewise, Gumienak et al. (2011) suggested undertaking a job review to distinguish critical/essential and frequently occurring tasks.

However, in some occupations, emergencies are not always of short duration; wild-land fires, in particular, can continue to blaze for weeks. In other occupations such as postal carrier (where physical ability has also been a matter of contention in setting required work-rates), vigorous activity may be required for 4-5 hours per day (Shephard, 1983), and in some jobs fatigue from using an excessive fraction of long-term aerobic power can pose dangers for the individual employee, coworkers, and the general public.

Task components

Perhaps the most contentious issue in worker assessment is the centrality of physical characteristics such as aerobic power and muscular strength to task performance. Coordination, balance and agility may reduce the physical demands of a given task substantially for some employees (Baechle and Earle, 2008). Moreover, some task analyses, such as that of Bard et al., (1985) have suggested that aerobic power and muscular strength are only required to perform a small fraction of a safety-officer's duties (although these commonly reflect emergency situatiuons, and are thus an important component of a screening protocol). Our analysis of performance among Inuit hunters living in the physically demanding conditions of the Canadian Arctic shows that the hunting success of arctic villagers is related more to intelligence and accumulated experience than to a large maximal aerobic power or muscular strength (Shephard and Rode, 1996). Although the instinctive reaction to a dangerous situation is the maximum deployment of aerobic power and muscular strength in a "fight or flight" reaction, a considered and less physical response may ultimately be more effective in achieving the desired objective. Comparisons between male and female police officers have demonstrated that in comparison with their male counterparts, the women resolve conflicts more frequently by non-physical means, thus compensating for lesser physical strength (Gaul and Wenger, 1992 ; Health Service, 1996). However, emergency workers must meet job demands from their first day of employment, and the emergency tasks in test batteries reflect issues that cannot be resolved by non-physical means; experience and approaches to a task are not criteria of current fitness screening. In the future, the ability to use intelligence in deploying available physical assets could possibly be assessed through a task simulation with several possible options for completion, although at present subject matter experts must attest that a test task could not be accomplished in a less demanding manner.

In discussions of public safety, another controversial topic is the possible inability of an employee to carry out a necessary task because of a medical emergency, particularly a myocardial infarction. The normal risk of such a contingency is increased several fold by the combination of intense physical activity and a dangerous situation (Shephard, 1981). It has been argued that the likelihood of future medical issues such as a heart attack should not be considered if a person is currently able to meet employment standards (Jamnik et al., 2013). There are occupations such as commercial airline pilot where the public is exposed to the individual's risk of a heart attack for quite long periods (for example, 600 flying hours per year), and exercise ECG evidence pointing to future risk of myocardial infarction is a sufficient cause for disqualification from employment (Joint Aviation Authorities, 2009). To some extent, this is an issue also for emergency workers, although Meorin has specified that markers of cardiac risk such as body mass index are medical issues, and cannot be included in a fitness test.

Certainly, there are pragmatic arguments against considering future health in many public safety occupations. Although a person with cardiovascular risk factors and an abnormal ECG has a several-fold increase in the risk of sustaining a heart attack relative to peers of the same age, if such criteria were to be applied to an entire labour force, a high proportion of those denied employment would be excluded because of "false-positive" test results (Shephard, 1981). Moreover, the risk of a heart attack in an apparently healthy individual of working age is sufficiently low that even if we could identify a sub-segment of the population with a 10-fold increase of risk, the likelihood of failure to complete a safety-related task of perhaps 15 minutes duration would remain extremely low (Shephard, 1991). Plainly, there is scope for a clearer definition of what is an acceptable level of risk.

In any event, employers are not absolved from the need to take all possible measures to enhance the cardiovascular fitness of their entire labour force, thus reducing the likelihood of such a catastrophe.

Test accuracy

Perhaps the largest concern when selecting workers is the question of test accuracy (Shephard and Bonneau, 2002). We will consider this issue from the viewpoint of aerobic performance, although the problem is likely to be at least as large for minimum standards of muscle strength and any other criteria that may be imposed. In terms of maximal oxygen intake, there are at least 3 potential sources of error- the prediction process, effects due to inter-individual differences of body mass, and intra-individual test-retest differences.

Prediction errors. When testing a large group of employees, the maximal aerobic power may be measured directly, or built into the test protocol, but it is commonly predicted from a field procedure such as a shuttle run or the Canadian Aerobic Fitness Test. Most prediction formulae have a standard error of 10-15% (Shephard, 1994), and it has been suggested that test scores should include an allowance for this error or a "borderline" category to allow for this issue (Reilly and Tipton, 2008). Employers must make a subjective decision between an "exclusive" standard with (theoretically) a 99.5% confidence of correct categorization, or an "inclusive" standard, with 87.5% correct categorization (Milligan et al., 2010; Tipton et al., 2013). Nevertheless, either of these choices carries substantial problems. In one sample of 100 individuals who undertook a simple step test, an "inclusive" standard passed 24 of 100 test candidates who should have failed, and placed a further 5 unsatisfactory employees in the borderline category, while the "exclusive" standard failed 42 of 100 candidates who should have passed, and placed a further 25 adequate individuals in the borderline category (Tipton et al., 2013).

The error of such predictions is further increased by an adverse environment (hot or cold conditions), and especially nervousness upon the part of the employee who is being tested (Shephard and Bonneau, 2002).

Effects of body mass. Minimum standards of aerobic power and muscular strength are commonly expressed in units of ml/[kg.min] and Newtons, respectively. However, it has long been recognized that the energy cost of most tasks varies with the individual's body mass (Brown, 1966 ; Godin and Shephard, 1973 ; Malhotra et al., 1962 ; Passmore and Durnin, 1955). Inter-individual differences in body mass thus have a substantial bearing upon a person's ability to undertake a given task, with heavy workers (whether muscular or obese) at a substantial disadvantage in meeting the prescribed target values for tasks demanding aerobic effort (Bilzon et al., 2001). In most tasks, all or much of the body mass must be displaced, and this may thus be a justifiable penalty to impose when developing the aerobic component of the test protocol. On the other hand, a lighter person is disadvantaged disproportionately, both in testing and on the job, by the weight of standard equipment (guns, belts, communications equipment, helmets and other protective clothing), and if the task involves carrying or dragging heavy loads, a muscular person is much better equipped to carry out the task than a lighter individual. Some aspects of this component of working ability can be assessed by testing performance during a backpack run (Vanderburgh and Flanagan, 2000) or the individual's fully-loaded performance on a test circuit (Mott, 2013).

A low body mass can also influence the accuracy of test scores. For instance, female candidates have been handicapped during push-pull tests because they have lacked sufficient body mass to avoid slipping on the smooth floor of a gymnasium (Gaul and Wenger, 1992).

Intra-individual test-retest errors. Jamnik and colleagues (2013) addressed the topic of test-retest errors in terms of the test-retest reliability of group scores for completion of a test circuit. They found similar mean values on successive days, and test-retest correlation coefficients that ranged from 0.75 to 0.98 (Jamnik et al., 2013). However, from the viewpoint of employment standards, the critical issue is consistency in the performance of the individual rather than that of the group. In the Canadian laboratory accreditation project, laboratories were required to measure maximal oxygen intake with a tolerance of 2.5%. However, if an individual repeats a maximal oxygen intake test in the laboratory over the course of several weeks, there is likely to be a 5% difference between successive determinations (Wright et al., 1978) and even with a careful standardization of protocol, the test-retest variation is likely to rise to 10-15% under field conditions (Shephard, 1991). Because of this issue, several formulations of the Canadian Aerobic Fitness Test (Shephard and Bouchard, 1993; Shephard et al., 1987) including that in current use limit evaluation of an individual to a five category scale (Shephard, 2014).

A 10-15% test-retest error seems uncomfortably close to the margin of perhaps 20% between physiological data for a satisfactory and an unsatisfactory employee, adding substantially to the probability of misclassification. Although times on a test circuit have greater face validity, such assessments are affected by even more variables than a direct or indirect measurement of maximal oxygen intake, and in consequence their scores are likely to have an even greater intra-individual variation.

Much of the test-retest error is random in distribution, and if an unsuccessful candidate is allowed to repeat the tests, there should be a progressive regression of data towards the true mean value (Shephard, 2003). However, if the error is largely random, acceptance of a higher score at a second attempt may simply replace a justified rejection by an unwarranted acceptance of a candidate.

The problem of limited test accuracy is not unique to industrial fitness assessments. In essence, almost all repeated physiological measurements have an error of at least 5%, even if determinations are made under carefully standardized laboratory conditions. The information obtained by a clinical physiologist is reasonably satisfactory when assessing the status of a group of subjects, but it becomes highly questionable if it is used to provide a unique rating of an individual's performance (Shephard and Turner, 1959).

It will be difficult to overcome this inherent difficulty in the context of employment standards, and at some stage, the inherent weakness of the numbers that are used is likely to be the subject of legal scrutiny. Until this issue has been regulated, there may be need to consider a tactic that I have previously suggested- the offering of a probationary appointment to those meeting accepted test criteria, with retention on the labour force being contingent upon satisfactory "on the job" performance (Shephard, 1991). Some have expressed the hope that unsuitable candidates would drop out of training programmes of their own volition over a period of rigorous preliminary training; unfortunately, experience does not support this suggestion (Nottrodt and Celentano, 1984).

Conclusions

In jobs where public safety is an issue, much effort has been expended to date in identifying tasks critical to successful job performance, and in setting objective and reliable tests of these abilities with both construct and face validity relative to the required employment (Shephard and Bonneau, 2002). Assessments have generally been safe to perform (Gaul and Wenger, 1992; Jamnik et al., 2013), and have been performed under realistic conditions in terms of clothing and equipment. Much thought has also been given to methods of accommodating sub-groups of the population such as women and smaller men who might otherwise have difficulty in meeting the required standards (Jamnik et al., 2013).

Group data for the proposed criteria can be obtained with reasonable accuracy, but the challenge (as with most physiological testing) is to make individual assessments with sufficient accuracy to warrant career-ending decisions for potential employees (Shephard, 1991 ; Shephard and Bonneau, 2002). It has yet to be demonstrated that the techniques currently make more than a very crude categorization as to whether the physical abilities of an individual employee will meet the emergency demands encountered on the job. One study estimated that a 45 second increase in the time that police officers were allowed to complete a 4 minute test circuit would allow 95% of criminals to escape (Farenholz and Rhodes, 1990). However, there also remains a need for a comprehensive study examining the risks to public safety associated with various levels of fitness, and a critical assessment of the gains in safety that are achieved by the implementation of employment standards.

Author's Qualifications

The author's qualifications are as follows: Roy J. Shephard, M.B.B.S.; M.D. [Lond.]; Ph. D.; D.P.E.; LL.D.

References

Åstrand, I. (1967). Degree of strain during building work as related to individual work capacity *Ergonomics 10,* 293-303.

Baechle, T., and Earle, R. W. (2008). *Essentials of strength training and conditioning* (3rd ed.). Champaign, IL: Human Kinetics.

Bard, C., Fleury, M., John, J., Lagassé, P., and Roy, B. (1985). *Elaboration des normes physiques d'admissionaux corps d'agents de la paix* Québec, QC: Facculty of Physical Education, Laval University.

Bilzon, J. L. J., Allsopp, A. J., and Tipton, M. J. (2001). Assessment of physical fitness for occupations encompassing load-carriage tasks *Occupational Medicine, 51*, 357-361.

Bink, B. (1962). The physical working capacity in relation to working time and age *Ergonomics 5,* 25-28.

Bonjer, F. H. (1968). Relationship between physical work capacity and allowable caloric expenditure In W. Rohmert (Ed.), *International Colloquium on musculoar exercise and training* (pp. 86-98). Darmstadt, Germany Gentner Verlag.

Brown, J. R. (1966). The metabolic cost of industrial activity in relation to weight *Medical Services Journal of Canada 22,* 262-272.

Clemes, S. A., Matchett, N., and Wane, S. L. (2008). Reactivity: an issue for short-term pedometer studies? *British Journal of Sports Medicine, 42*, 68-70.

Costill, D. L. (1972). Physiologfy of marathon running *Journal of the American Medical Association 221,* 1024-1029.

Farenholtz, D. W., and Rhodes, E. C. (1990). Normes canadiennes pour la capacité physique des policiers. *Journal du Collège Canadien de Police 14 (1)*, 39-50.

Farenholz, D. W., and Rhodes, E. C. (1990). Recommended Canadian standards for police physical abilities *Canadian Police College Journal 14 (1),* 37-39.

Gaul, C. A., and Wenger, H. A. (1992). *RCMP Physical abilities requirement evaluation demonstration project.* School of Physical Education, University of Victoria, Victoria, BC.

Gledhill, N., and Shaw, C. (1996). *Report on the characterization, test construction and validation phase of the medical, physical, skills and abilities project* Toronto, ON: Race Relations and Policing Unit of the Ministry of the Solicitor General and Correctional Services, (l), 104. (PREP).

Godin, G., and Shephard, R. J. (1973). Body weight and the energy cost of activity. *Archives of Environmental Health 27,* 289-293.

Gumienak, R., Jamnik, V., and Gledhill, N. (2011). Physical fitness *bona fide* occupational requirements for safety related physically demanding occupations; development considerations. *Health & Fitness jopurnal of Canada 4 (2)*, 47-52.

Health Service Directorate. (1996). *PARE Standard Summary* Ottawa, ON: Royal Canadian Mounted Police.

Jamnik, V., Gumienak, R., and Gledhill, N. (2013). Developing legally defensible physiological employment standards for prominent physically demanding public safety occupations: a Canadian perspective. *European Joiurnal of Applied Physiology 113,* 2447-2457.

Joint Aviation Authorities. (2009). *Civil Aviation Medicine*. Englewood, CO: Global Engineering Documents.

Lagassé, P., Bard, C., Fleury, M., Jobin, J., and Roy, B. (1985). *Elaboration des normes physiques d'admission aux corps d'agents de la paix*, Québec, QC: Département d'éducation physique de l'université Laval.

Louhevaara, V., Smolander, Z., Korhonen, O., and Tuomi, T. (1986). Maximal working times with a self-contained breathing apparatus. *Ergonomics 29,* 77-85.

Malhotra, M. S., Ramaswamy, S. S., and Ray, S. N. (1962). Influence of body weight on energy expenditure. *Journal of Applied Physiology 17*, 433-435

Milligan, G. S., House, J. R., Long, G., and Tipton, M. (2010). *A recommended fitness standard for the oil and gas industry*. London, UK Energy Institute Report. ISBN 9780-8529-3562-0.

Mott, B. (2013). *Contemporary physical employment standards for fire and rescue NSW* Paper presented at the Shaping tomorrow together conference, Melbourne, Victoria.

Nottrodt, J. W., and Celentano, E. J. (1984). Use of validity measures in the selection of physical screening tests. Paper C-04 *International Conference on Human Ergonomics*. Toronto, ON: Human Factors Association of Canada.

Passmore, R., and Durnin, J. V. G. A. (1955). Human energy expenditure. *Physiological Reviews 35,* 801-840.

Reilly, T., and Tipton, M. (2008). Chapter 12. Fitness requirements for offshore rescue. In C. J. Brooks (Ed.), *Survival at sea for marine aviators*: NATO Research & Technology Organization AGARDograph-HFM-152.

Rhodes, E. C., and Farenholtz, D. W. (1992). Police Officer's Physical Abilities Test compared to measures of physical fitness. *Canadian Journal of Sport Sciences 17,* 228-233.

Shephard, R. J. (1981). *Ischemic heart disease and exercise*. London, UK: Croom Helm.

Shephard, R. J. (1983). The workload of the postal carrier. *Journal of Human Ecology 11,* 151-164.

Shephard, R. J. (1991). Occupational demand and human rights: public safety officers and cardiorespiratpry fitness. *Sports Medicine 12,* i94-109.

Shephard, R. J. (1994). *Aerobic fitness & health*. Champaign, IL Human Kinetics.

Shephard, R. J. (2003). Regression to the mean: A threat to exercise science? *Sports Medicine 33,* 575-584.

Shephard, R. J. (2014). Certified exercise physiologists and exercise prescription: Evolution of the PAR-Q and Canadian Aerobic Fitness Test. *Journal of Physical Activity & Health, In press*.

Shephard, R. J., and Bonneau, J. (2002). Assuring gender equity in recruitment standards for police officers. *Canadian Journal ofApplied Physiology 27,* 263-295.

Shephard, R. J., and Bouchard, C. (1993). A new approach to the interpretation of Canadian Home Fitness Test scores *Canadian Journal of Applied Physiology, 18*, 304-316.

Shephard, R. J., Ferris, B., Craig, C., and Haines, D. (1987). Predicting the cardiovascular fitness of Canadians *Canadian Journal of Sport Sciences, 12,* 152-155.

Shephard, R. J., and Rode, A. (1996). *The health consequences of "modernization"*. London, UK: Cambridge University Press.

Shephard, R. J., and Turner, M. E. (1959). On the probability of correct diagnosis by pulmonary function tests. *Thorax 14,* 300-304.

Speechly, D. P., Taylor, S. R., and Rogers, G. G (1996). Differences in ultra-endurance exercise performance - matched male and female runners. *Medicine & Science in Sports & Exercise 28,* 359-365.

Tipton, M. J., Milligan, G. S., and Reilly, T. J. (2013). Physiological employment standards. 1. Occupational fitness standards: objectively subjective? *European Journal of Applied Physiology 113,* 2435-2446.

Tschakovsky, M. E., and Hughson, R. L. (1999). Interaction of factors determining oxygen uptake at the onset of exercise. *Journal of Applied Physiology 86,* 1101-1113.

Vanderburgh, P., and Flanagan, S. (2000). The backpack run test: a model for a fair and occupationally relevant military fitness test. *Military Medicine 165,* 418-421.

Wright, G. R., Sidney, K. H., and Shephard, R. J. (1978). Variance of direct and indirect measurements of aerobic power. *Journal of Sports Medicine & Physical Fitness 18,* 39-42.