AD-P003 257

PHYSICAL PERFORMANCE TESTS AS PREDICTORS OF TASK PERFORMARCE

T.L. Doolittle Ph.D.
University of Washington

Oscar L. Spurlin Ph.D., and M. Peter Scontrino Ph.D.
Ergometrics and Applied Personnel Research Inc.

Abstract of Panel Presentation

Test Development:

The more arduous the task, the greater the intensity of force which must be applied per unit of time to overcome resistance or achieve rate. Intensity is commonly called "workload" with magnitude expressed in appropriate units of power. Two complex factors determine the limits for which an individual can produce energy and generate the requisite power:

1. capacity to utilize oxygen, and
2. ability to generate muscular tension.

The former is called "aerobic power" and the latter "strength". For repetitive tasks, over an 8 hour workday (with normal breaks), individuals normally will not function at an intensity greater than 40% of their maximum aerobic power and/or 15% of maximum strength. The limit for occasional lifting, while influenced by posture and body mechanics, normally should not exceed 70% and never exceed 90% of the individual's maximum strength.

Aerobic power requirements are estimated from heart rate response to known and unknown workloads. From rest to approximately 70-80 percent of maximum aerobic power, increases in heart rate correlate linearly with increases in energy expenditure and oxygen consumption (Figure 1).

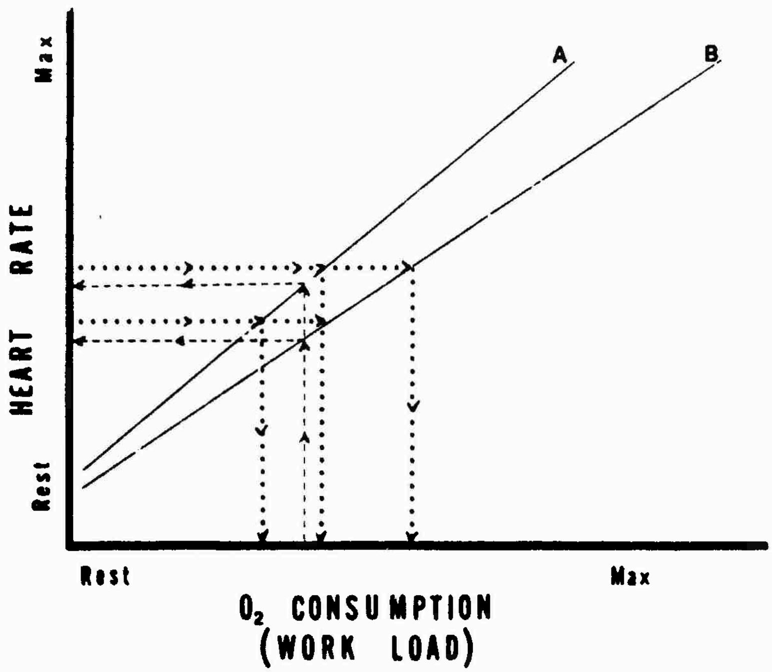
Further, since the heart can beat only so fast and still function as an effective pump, an individual's maximum functional capacity can be projected from the heart rate response at a given sub-maximum workload. Inter-individual comparisons also are possible. In Figure 1, a given workload (dashed line) is relatively more demanding for person "A" than for person "B", and since both will "max"at about the same heart rate, "B" has more reserve and should be capable of a higher maximum workload. Once the baseline relationship between heart rate and workload has been established for each individual, heart rate on job tasks (dotted lines in Figure 1) enables one to accurately estimate the metabolic requirements. This protocol has been utilized and is widely reported for a variety of occupational endeavors in the literature (Astrand, 1977).

Strength requirements are established by determining the forces required to lift, push, pull, shovel, slide, etc. as appropriate. Multiple samples are necessary to account for variability. Heights and distances the objects must be moved, and total amount of material per unit of time must be obtained. From these data strength requirements are determined and expressed in terms of pounds per lift (for occasional endeavors) or foot pounds per minute for repetitive tasks. While the aerobic power requirements can be generalized for measurement on a treadmill, cycle ergometer, or bench step, strength testing should closely replicate the task to insure that the requisite muscle groups are being tested.

For occasional lifting, requiring the individual to lift an object of similar weight five (5) times will insure that it does not exceed approximately 85% of capacity. For repetitive strength tasks 67% of capacity can be performed at 20 contractions per minute for short durations. Thus, for a workload requirement of 750 ft-lb/min, there are two defensible alternatives:

- a single maximum test requiring 6250 ft-lb/min, or
- a three to five minute effort at 4200 ft-lb/min, accomplished at a rate of 20 repetitions per minute.

FIGURE-1 Heart Rate & Work Load



Validity Evidence and Impact on Protected Groups.

From the foregoing discussion it can be seen that it is possible to replicate the significant components of physically demanding occupations. If a test can be demonstrated to represent important job components it is valid to use the test in applications such as pre-employment screening. Nevertheless, because of the legal guidelines and changing professional standards surrounding test validation, there are some important issues to consider in order to firmly establish the defensibility of a physical performance test.

Content vs. criterion-related validity

The primary issue here is whether job analysis alone can provide a sufficient basis for validation. Our experience is that it is, in fact, a superior method. Job analysis procedures as described above, provide a detailed description of actual task requirements. This description is independent of the characteristics of current incumbents, it is not dependent on large numbers of employees performing a particular job, and it is not dependent upon supervisory ratings or other highly subjective methods.

The difficulty with criterion-related validation studies is the criterion. Physical performance tests are often validated through use of work samples as performance criteria. Examples can be given to show that the subjectivity and potential for measurement error is greater in this process, than if the entire research effort was concentrated on quantifying actual work tasks and replicating the critical components of these activities.

An issue for both these approaches is which tasks are truly critical to job performance. Is a task that is performed twice a year a sufficient justification for a screening standard? Emergency service personnel are often screened on the basis of worst-case scenarios. This type of rationale requires substantially more justification. Another frequently confronted case is the 'labor pool' where only a certain proportion of individuals are ever assigned the arduous tasks.

Simulation

Perhaps one the more difficult questions concerns the degree of 'true-to-life' simulation necessary to be assured of a job-related test. That is, does a test need to involve exactly the same muscle groups used in exactly the same range of motion to provide an accurate prediction of job success.

In cases where the objective is to select among a relatively large number of applicants for a given opening, a high degree of precision is usually not necessary. This finding results from the moderate to high correlation between different types of physical performance tests. Leg strength is different from trunk strength, and stamina is different from both; yet tests of these dimensions might correlate on the order of $r=.40$. Thus, a typical finding would be that a single test can often provide roughly the same rank ordering of candidates as an entire battery of tests.

In situations that involve the use of highly developed physical abilities, then simulations are more appropriate in order to assess specific job-related capabilities that it would be impractical to learn or develop on the job. Rehabilitation therapy is another instance where more detailed knowledge of specific physical abilities may be needed. We have seen a real need for return-to-work standards that are based on objective job related measures of capability.

Impact on protected groups.

Physical performance tests will have an impact on women and to a lesser extent other minorities of smaller average stature. There are, of course real population differences in strength and endurance, and tests will reject a higher proportion of females than males. This does not mean that physical performance tests are automatically subject to successful legal challenges.

It is reasonable and legal to require job applicants to possess the qualifications necessary to perform the job effectively and safely. Physical performance tests will be well accepted by all concerned, provided:

- Job-relatedness is shown through research that conforms to legal and professional standards

- There are no alternative tests or procedures which could identify qualified candidates without the same impact.

- The job itself cannot be easily redesigned to eliminate the physically limiting tasks.

- The tests are used to screen 'in' physically qualified females and minorities.

In practical terms, a hiring agency simply needs to recruit and screen more female applicants to insure the placement of enough qualified and potentially successful individuals.

Utility: The Bottom Line

Over thirty years ago Brogden (1949) used linear regression to demonstrate the relationships of cost of selection, validity, and the selection ratio to the utility of a test. In this context utility refers to the dollar savings to an organization resulting from the higher performance (improved productivity) of those employees selected using a validated test. The higher the average test score of those selected, the greater the utility of the test.

Recently Schmidt et. al. (1979, 1982, 1983) and Cascio (1982) have revived interest in utility theory by providing inexpensive and straightforward procedures for estimating the utility of a selection procedure. Cascio (1982) has extended this process to assessing the utility of a wide range of human resource programs in organizations.

The basic formula for estimating the utility (the total improvement in performance over random selection) of a test as follows:

$$\Delta\mu = N_s r_{xy} SD_y \lambda / \phi - N_s c_y / \phi \qquad (1)$$

where:

$\Delta\mu$ = total gain over random selection
N_s = number of applicants selected
r_{xy} = validity of the predictor, when evaluated against the dollar-valued job performance criterion
SD_y = standard deviation of the dollar-valued job performance criterion
λ = the ordinate (height) of the normal curve corresponding to the predictor cutoff (the selection ratio)
ϕ = the selection ratio
c_y = cost of putting one person through the selection process

If one wishes to compute the gain in utility from substituting one selection procedure for another, the formula is:

$$\Delta\mu = N_s(r_1 - r_2)SD_y\lambda/\phi - N_s(c_1 - c_2)/\phi \qquad (2)$$

Where all terms are as defined above, except that r_1= validity of the new procedure, C_1= cost of the new procedure, C_2= cost of the old procedure.

In these formulas the one piece of information that is not readily available is the SD_y. Both Schmidt and Cascio present simple methods to estimate this dollar valued variance in performance.

Of course, physical performance testing yields an equal or larger benefit in terms of cost savings resulting from decreased lost time injuries. Using the above procedure and including the safety related benefits has led us to conclude that physical performance testing can lead to an increase in value of 15-30% for each individual hired. This translates into $2000 - $3000 per employee annually in a typical application.

REFERENCES

Astrand, P.O., & Rodahl, K. Textbook of Work Physiology (2nd Ed.). McGraw-Hill, New York, 1977.

Brogden, H.E. When testing pays off. Personnel Psychology, 1949, 2, 171-183.

Cascio, W.F. Applied Psychology in Personnel Management. Reston, VA: Reston Publishing Company, 1982.

Schmidt, F.L., Hunter, J.E., McKenzie, R.C., & Muldrow, T.W. Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 1979, 64, 609-626.

Schmidt, F.L., Hunter, J.E., & Pearlman, K. Assessing the economic impact of personnel programs on workforce productivity. Personnel Psychology, 1982, 35, 333-347.

Schmidt, F.L., Hunter, J.E. Quantifying the economic impact of psychological programs on workforce productivity. Amer. Psychologist, 1983, 38, 473-478