



PHYSICAL ABILITY TESTING

BACKGROUND

A test of physical constructs assesses a candidate's ability to perform essential physical tasks of the job. Two major types of physical tests most often used by public safety agencies are as follows: physical fitness tests and job-task simulation tests. Each type will be discussed, and information on their validation and legal defensibility will be presented.

PHYSICAL FITNESS TESTS

Physical fitness tests require candidates to perform exercises (e.g., push-ups, pull-ups, sit-ups, timed run, etc.) that are designed to assess underlying physical fitness levels. It is assumed that if a candidate possesses a high enough level of fitness, he/she will be able to manage the physical aspects of the job. The major disadvantage of this type of test is that it does not appear to be job-related. For example, doing push-ups or jumping to a certain height are not tasks that must be completed while on the job. Thus, the physical fitness test is an indirect measure of a candidate's ability to perform essential physical task requirements of the job. Each agency must go through the rigorous undertaking of conducting both construct and criterion validation studies in order to choose the exercises as well as set the standards for passing.

JOB-TASK SIMULATION TESTS

Job-task simulation tests require candidates to perform tasks in a timed course that are replications or simulations of actual physical job

tasks. The types of tasks that may be included in job-task simulation tests for law enforcement agencies include the following: dummy drag, stair climb, fence climb, etc. The types of tasks that may be included in job-task simulation tests for firefighting agencies include the following: ladder raise, ladder heel, hose advance, dummy drag, stair climb, equipment carry, forcible-entry simulation, ceiling breach and pull simulation, etc. The physical job task replications or simulations that are chosen for inclusion are representative of essential physical job tasks based on results from a thorough physical task job analysis. Each of the tasks included must be performed individually by the candidate without any assistance. Also, the tasks chosen for inclusion should not require the candidate to have any prior experience or training in order to complete the task. Safety is another concern that is considered when including any of the test components. Job-task simulation tests operate on an absolute standard. Each candidate must complete each task in the course and complete the entire course by the established cutoff time in order to pass the test.

SETTING CUT SCORES FOR JOB-TASK SIMULATION TESTS

One of the key decisions that must be made when using a job-task simulation test is determining the standard for passing. After the job-task simulation test has been developed for an agency, the next major step is to conduct a field test with incumbents. The sample of incumbents selected to take part in the field test should be diverse and contain an oversampling

of women and minorities. If at all possible, the incumbents should be randomly selected from the relevant ranks for inclusion in the field test sample. The incumbents must be told to give their best effort when participating in the field test and to take performance on the test seriously. Each incumbent would then go through the job-task simulation test, and their completion time would be recorded. The incumbents' time data from the field test would be used to determine the passing cutoff time.

The cutoff time that is established must be reasonable and consistent with the performance of qualified incumbents. There are many methods agencies can use to set the passing score. Whatever method that is chosen must be documented with sufficient detail on the approach taken in setting a passing cutoff time. The method that I/O Solutions recommends in setting the cutoff time for such a test is to use a norm-referenced approach. The candidates who complete the job-task simulation test by the cutoff time will pass the test, while candidates who are not able to complete the test by the cutoff time will fail the test. When using this approach, the data from the field test is used to obtain the average and the standard deviation across all incumbents' completion times for the test. The passing time that is usually recommended for agencies to use is two standard deviations below the average of the times obtained in the field test. A cutoff time that is two standard deviations below the average will represent a point at which the vast majority (approximately 98 percent) of all qualified incumbents in a normal distribution would pass the test.

VALIDATING PHYSICAL TESTS

A test is said to have validity when it is supported by sound evidence so that the inferences drawn from the actual test scores are appropriate and meaningful. Due to the strict scrutiny of physical

tests in the legal arena, agencies need to ensure that any physical test that is given adheres to the Uniform Guidelines on Employee Selection Procedures. The Uniform Guidelines present three types of validation strategies that may be used to support the use of a test: content, construct and criterion. Content validation accumulates evidence to make the assertion that the test incorporates content that is representative and relevant to the job. Content validation can be used to validate job-task simulation tests but cannot be used to validate physical fitness tests. Construct validation accumulates evidence to support that the test is measuring the underlying construct that it is intending to assess (e.g., fitness, strength, endurance, etc.). Construct validation, however, requires extensive effort as it usually involves a series of research studies, which includes criterion related validity studies and which may include content validity studies. Construct validation would be most relevant for validating physical fitness tests but could not be used to validate their standards for passing. Criterion validation accumulates evidence to support the relationship between performance on the test and later job performance. Criterion validation can be used to validate job-task simulation tests, physical fitness tests and their standards for passing. The Uniform Guidelines state that one or more of these types of validity evidence must be demonstrated when adverse impact exists.

RECOMMENDATIONS WHEN VALIDATING PHYSICAL TESTS

Sample Composition and Size: First and foremost, when conducting a criterion validation, it is imperative to use a diverse sample of the agency's incumbents to go through the physical test. Women and minorities need to be included in the sample; otherwise the legal defensibility of the physical test is threatened (e.g., *United States of America*

v. City of Erie, 2005). Also, incumbents from various age groups should be included. It is recommended that the incumbents be randomly selected for inclusion in the sample. The sample size should be sufficiently large, no smaller than 30, and preferably larger for criterion validation studies (Biddle & Sill, 1999).

Validating the Physical Test as a Whole vs. Individual Components: Another issue that has been raised by the courts is to be consistent in how the test is administered and scored and how the test is validated (e.g., *United States of America v. City of Erie*, 2005). For example, if a job-task simulation test has six different components that the candidates must complete in four minutes or less in order to pass, then the test must be validated as a whole and not merely by its parts. Thus, agencies need to be consistent when conducting criterion validation studies to be sure that the test is validated in a manner consistent with how it will be administered.

Job Analysis: If an agency wishes to utilize a content validation approach to validate its job-task simulation test, it must show that a proper and current job analysis was conducted (e.g., *Legault v. Russo*, 1994). The job analysis that supports a job-task simulation test must be specific in the physical tasks required for the job in question. The job analysis data must include ratings from its subject-matter experts on importance, frequency and level of proficiency required for each of the tasks. To the extent that the job analysis is specific in providing this aforementioned information, the agency will have strong evidence demonstrating the job-task simulation test's content validity. This job analysis evidence is crucial if the job-task simulation test is ever legally challenged. It is then not unexpected that the courts do not accept anecdotal evidence in place of job analysis data. Assertions that the job-task simulation test includes components that are

similar to the actual job are without legal merit and are not sufficient evidence to support the job-task simulation test's content validity.

Transportability Evidence: When an agency wishes to use the same physical test as one conducted by another agency, it must gather evidence to support its transportability. The Uniform Guidelines allow for agencies to transport validity evidence from another agency if the following conditions have been met. First, the agency that has developed the physical test must have sound criterion validation evidence as well as a proper job analysis to support its use. Second, the agency wishing to borrow the physical test must conduct a proper job analysis and must demonstrate adequate similarity with the agency it is borrowing the physical test and validity evidence from. Third, evidence of fairness (e.g., adverse impact against women) must be collected. If an agency borrows a physical test from another agency, it should go through the accepted process of obtaining the required evidence to support its use. Courts do not side with agencies that copy what other agencies use for their physical tests, especially with no support of a thorough job analysis and criterion validation study (e.g., *Legault v. Russo*, 1994).

ENHANCING THE LEGAL DEFENSIBILITY OF PHYSICAL TESTS

Meeting the Job-Related and Consistent with Business Necessity Standard: Any selection test that is implemented in an agency must be job-related and consistent with business necessity as dictated by Title VII of the Civil Rights Act of 1964 as well as the Americans with Disabilities Act of 1990. If the physical test is ever challenged in court, the agency bears the burden of proving that the test is job-related and consistent with business necessity. Agencies wishing to demonstrate that their physical test

is job-related should present how the test approximates tasks, skills and abilities required of the job at a level that is expected of job incumbents, citing job analysis information along with validation studies. To demonstrate that the physical test is consistent with business necessity, agencies need to present evidence that performance on the test predicts later performance on the job. Thus, an agency meets its burden when it presents proper evidence that its physical test closely approximates and effectively measures tasks, skills and abilities that are important to success on the job (Hollar, 2000).

Adverse Impact and Physical Tests: In the context of using physical tests for selection, adverse impact occurs when an identical standard for passing is applied to everyone despite the fact that it leads to a substantial difference in selection for members of a particular group (e.g., women). The federal government uses the four-fifths rule to determine if adverse impact occurs. If any protected group has less than four-fifths of the selection rate of the group with the highest selection rate, then evidence of adverse impact exists. Most physical tests have been accepted to have adverse impact against women. Due to immutable physiological differences between men and women, women fail physical tests at a much higher rate than men. The issue then becomes one that forces the agency to defend its physical test as being job-related and consistent with business necessity. If an agency can demonstrate that its physical test is supported by sound evidence demonstrating that it is job-related and consistent with business necessity, then it can continue to use that test. The agency should also consider whether alternative tests exist that offer the same benefits of the physical test in selecting candidates who will be successful in the job but without the adverse impact. If an agency has

diligently developed its physical test using a proper job analysis, supported its use with appropriate validation studies and tested for fairness in its applicant pool, then the physical test can be confidently defended if it is ever challenged.

CONCLUDING REMARKS

When developed and used correctly, a physical test can be an important selection tool for public safety agencies. Due to the physical nature of several positions within public safety agencies, a physical test is thought of as an important tool to ensure that the candidates selected for these positions are able to perform all of the essential tasks of the job. There is always a legal risk with using physical tests as a selection tool since there is inherent adverse impact against women due to immutable physiological differences. Nonetheless, an agency can take the required steps when developing and validating its physical test that will allow them to confidently defend their tool if it is ever challenged.

REFERENCES

- Biddle, D. & Sill, N. (1999). "Protective Service Physical Ability Tests: Establishing Pass/Fail, Ranking, and Banding Procedures." *Public Personnel Management*, 28(2), 217-225.
- Equal Employment Opportunity Commission. (1978). *Uniform Guidelines on Employee Selection Procedure*; 43 FR 38295; 29 CFR Part 1607.
- Hollar, D. (2000). "Physical Ability Tests and Title VII." *University of Chicago Law Review*, 67, 777.
- Legault v. Russo*, 842 F. Supp. 1479, 1488 (D. N.H., 1994).
- United States of America v. City of Erie, PA*, 411 F. Supp. 2d 524 (W.D. Pa., 2005).