



FITNESS TESTS, STANDARDS, AND NORMS:

What is Valid? What is Legal?

There is much confusion over the development and use of physical tests, norms and standards within law enforcement. Terms have been misunderstood and the many legal issues have often been misinterpreted. Fitness tests and test scores used as standards without proper validation have been misapplied. There is also some controversy over whether job task simulation or fitness tests should be used.

FitForce™ and its associates (Fitness Intervention Technologies and Hoffman and Associates) reviewed many validation studies to define fitness standards, relevant legislation, recent court cases and the practical experience of those who have implemented fitness standards to clarify these issues.

### TESTING TERMS

- A **job task simulation test** is a test that measures the ability to perform a specific physical task of the job such as pushing a car, jumping a fence etc.
- A **physical fitness test** measures an underlying physical factor or “ability” that determines the capability to perform strenuous job tasks.
- A **physical abilities test** is the same as a fitness test. However, that term has been misused to also describe a job task simulation test.
- An **agility test** measures a specific fitness factor or motor skill called agility (the ability to move quickly and change directions). However, that term has been misused to also describe a job task simulation test.
- **Norm or percentile score** is a score point for given test that is based on a norm profile of a group of people tested. It is not based on any kind of job relatedness i.e. the prediction of the ability to perform essential functions of the job.
- A **standard** is a score on a test that has been determined to measure the minimal amount of the factor measured by the test (a fitness factor or performing a job task) that is required to do the job.
- **Job relatedness** of a test and test score (standard) means that meeting the test standard is predictive of being able to perform the essential functions of the job (job tasks). If a test measures a factor that has no to relationship to performing the essential functions of the job (job tasks) or if meeting the standard does not predict the ability to perform the essential functions of the job (job tasks) it is not job related. **Both the test and the test score used as a standard must show job relatedness.**

### ISSUES REGARDING LEGAL REQUIREMENTS

Public safety agencies have the discretion to use any fitness test, standard or program provided certain legal requirements are met. These requirements are set out by historical legislation. While there may be many different legal requirements, there are three that are of particular concern for using fitness tests, standards and programs:

## 1. **JOB RELATEDNESS – Civil Rights Act of 1964 and the Americans with Disabilities Act (ADA) of 1991.**

Both of these Acts requires that for an agency to use a test as:

- an employment standard for hiring
- for completion of training
- for incumbent retention
- for incumbent promotion

it must be **job related**. This means that the test score must be predictive of the ability to perform the essential functions of the job and be of business necessity. The test standard must be able to predict those who can do the job and those who can not.

## 2. **SAME JOB - SAME STANDARD – Civil Rights Act of 1991 – Section 106**

"It shall be unlawful employment practice for a respondent, in the connection with the selection of referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex or national origin."

### **This Act refers to applicant, recruit and incumbent standards.**

However, during the 1990's there was a faulty misinterpretation of this law by some that suggested that if you called a fitness test score as a fitness standard – not a job standard – then age and gender based norms could be used for applicant standards, but not for recruits or incumbents (since they would have to meet a job related standard). This was against the simple language of Section 106 and has since been deemed illegitimate. The current administration of the DOJ is of the opinion that an agency can not have a fitness standard for "fitness sake". It must be an employment standard that predicts the ability to perform essential functions. Only validated absolute fitness scores will meet that requirement.

## 3. **ADVERSE IMPACT – Civil Rights Act of 1964 – Section VII**

Section VII requires that an agency can not use a test standard that demonstrates adverse or disparate impact against a protected class (female, racial minority) unless the test standard is "job related". Adverse impact is defined as a protected class passing a test at a less than 80% rate of white males. **The key wording is that standards that show adverse impact can be used if there is data to show that the test standard is job related.** A recent court case (Lanning vs. SEPTA, U.S. District Court of SE Penn 2000) clearly affirmed this.

**Any standard is acceptable for being job related if it has been validated through methods** put forth by the EEOC in the Uniform Guidelines (U.S. DOJ, EEOC, Government Printing Office, 1978). In essence, an agency must meet a

strong burden of proof to document that the standard is predictive of the ability to do the job. If a standard has that validation data it is supported - if not it is not supported. The following summarizes the current state of this issue:

- Single standards that have an undergone acceptable validation process meet the legal requirements with a strong burden of proof behind them. They predict who can and who can perform the physical demands of the job.
- Age and gender based standards do not have any job related data to support their use. They do not predict who can and who can not do the job or can and who cannot complete training.

### **ISSUES REGARDING TEST VALIDITY REQUIREMENTS**

The term validity refers to the “soundness” and legitimacy of a test to accurately measure what it is supposed to measure and to be predictive of what the test is supposed to be used for. Validity also is a term to describe the process by which an agency goes through to define a test standard. The Uniform Guidelines point to four acceptable methods to conduct a fitness test validation:

- **Content validation.** The test measures the actual job-task and the program teaches that task. Physical fitness or ability tests cannot be validated with this approach. Only job task simulation tests can be validated with this method.
- **Construct validation.** The test measures an underlying fitness factor for performing job-task(s) and the program trains that factor. Fitness tests can be validated with this approach but not fitness standards
- **Criterion validation.** The test measures a predictive fitness factor for performing job-tasks and the program trains that factor. This is the preferred method for validating fitness tests and standards
- **Transportability study.** The Uniform Guidelines note that an agency can apply standards validated (by a content, criterion or construct methodology) from another agency if a commonality of job functions and tasks can be established. This is usually performed by comparing equivalent job task analysis data. Comparisons of job descriptions are not enough.

### **ISSUES REGARDING THE USE OF COOPER FITNESS TESTS AND NORMS**

The Cooper Institute has been involved in public safety fitness programming since 1979. The so called “Cooper test” that the Institute teaches in their Fitness Instructor courses has been and is being misapplied by law enforcement agencies. Recent legislation, litigation and validation studies have shown considerable misperceptions regarding the “Cooper “tests and associated norms that require clarification.

#### **The Cooper fitness test battery**

The tests that the Cooper Institute uses were not developed by the Cooper Institute. They are field based tests that exercise science has been using for years. The original battery

was put together in 1979 as a health related fitness battery to measure the major health related fitness areas. It consisted of the following:

1.5 mile run  
Sit and reach  
1 Minute sit up  
1 Minute full body push up for men and modified push up for women  
Body fat  
1RM bench press  
1RM leg press

Patients who come to Cooper Clinic for medical exams took those tests with the exception of the 1.5 mile run test. The patients took a treadmill test and a 1.5 mile run time was interpolated from it. Consequently, the 1.5 mile run norms are not true norms but interpolated numbers. Norms were created for the other tests based on that patient sample (approximately 50,000 people). There were no law enforcement officers in that sample. The norms were classified by age and gender since that is the normal classification for "health norms" The age and gender norms were also collapsed into "single "norms for comparison purposes. That basic test battery was then recommended for use by law enforcement agencies. The reasons that the Cooper norms had been applied in the past (prior to 1991) for law enforcement settings were as follows:

- Fitness was recognized as being important for health and physical performance and was justified as a healthy lifestyle area to prevent disability. Fitness was important for doing the job it just wasn't known then how much?
- In the absence of data documenting the specific amount of fitness necessary to do the job it was concluded reasonable to expect officers to be as fit as their age and gender norm group. Consequently the Copper age and gender based norms were recommended. The same %tile rank for each group was recommended to be applied as a standard, however the specific raw score for each %tile rank varied by age and gender group.

Based on the validation studies of FitForce and its associates the 300 meter run, agility run, and vertical jump were added to the battery for law enforcement agencies and the 1RM leg press was deleted because of logistical problems.

**The Cooper Institute does not conduct validation studies and has not validated the test battery or norms for being job related.** However, studies conducted by FitForce have shown those test items to be "job related". In other words, those tests measure the underlying physical abilities (fitness) to perform essential job tasks. The "Cooper" norms, however, have not been shown or have not been validated to be job related for law enforcement officers.

## Cooper norms

The use of the Cooper norms as standards all changed in the 1990's due to legal and test validity developments as follows:

- Research methodologies were developed and studies have been performed that can determine the amount of fitness necessary to perform strenuous physical tasks.
- Absolute single standards predict who can and who can not perform the strenuous physical tasks of the job. Age and gender based %tile norms do not and have no bearing on doing the job. They are totally invalid.
- Since single absolute standards predict the ability to do the job and the age and gender norms do not, scientific validity dictates the use of validated single standards.
- Since the literal interpretation of Civil Rights Act of 1991 prohibits the use of age and gender based standards it makes the most legal sense to use single absolute standards that have been validated.
- Since the ADA emphasizes that all employment standards must be job related, criterion validated single absolute standards are in compliance (because they predict job task performance) and age and gender standards are not (because they do not predict job task performance).
- Since the most rational way to conceptualize the job is as "Same job- Same standard", single standards make the most logical sense.

As mentioned, the Cooper norms are but profiles of individuals who have taken the tests. **They are not standards.** They are not job related standards in that the percentile scores do not predict the ability of officers to perform essential training or job tasks. Consequently, they should not be used for hiring officers, as graduation standards for completing an Academy or for incumbent retention.

However, the Cooper norms do have uses. They provide the percentile scores that can be used as goals for voluntary fitness programs that do not require mandatory participation or the meeting of a standard.

## **FITNESS TEST AND STANDARDS THAT ARE VALID AND LEGALLY DEFENSIBLE**

Due to litigation requirements, each agency or state POST must conduct a validation process to define fitness standards for itself. Specifically, this is a content, criterion or constructs validation or transportability study from agencies that have had such validation studies performed. Some core elements of such a validation consist of the following:

- Job task analysis (JTA) to define the essential physical tasks.
- Testing of stratified random samples of officers on criterion physical tasks and fitness tests.

- Statistical analysis to define those fitness score cut points that predict who can and who can not perform the physical job tasks.
- Standards implementation process that addresses leadership, administrative and program procedures.

### **Summary of validation studies**

FitForce and its associates have conducted validation studies on over 180 federal, state and municipal agencies. The summary of those validation studies reported in Police Chief (Collingwood, Hoffman and Smith, March 2004. p. 32-37) presents the following conclusions.

#### **Fitness areas that are predictive of performing physical tasks**

<b>Job tasks</b>	<b>Fitness areas</b>
Short pursuits	anaerobic power, leg power, agility
Sustained pursuit	aerobic power, agility, muscular endurance
Lifting/carrying	upper body strength, muscular endurance, agility, leg power
Jumping/vaulting	leg power, anaerobic power
Climbing	anaerobic power, aerobic power, muscular endurance, agility, upper body strength
Dragging/pulling	upper body strength, leg power
Pushing	upper body strength, muscular endurance, leg power
Dodging	agility, anaerobic power
Use of force < 2 minutes	anaerobic power, upper body strength, agility, muscular endurance
Use of force > 2 minutes	aerobic power, upper body strength, agility, muscular endurance
Bending/reaching	flexibility

#### **Fitness tests that measure the underlying fitness areas**

<b>Fitness factor</b>	<b>Test</b>
Absolute strength	1 RM Bench press raw score (lbs) and ratio score (weight pushed divided by body weight)

Explosive strength (leg power)	Vertical jump in inches
Dynamic strength (muscular endurance)	1 minute sit up (n) Maximum push up (n)
Extent flexibility	Sit and reach in inches
Aerobic power	1.5 mile run in min:sec
Anaerobic power	300 meter run in seconds
Gross Coordination (agility)	Illinois agility test in seconds

Each agency had specific job related fitness standards defined for that agency. There was a spread of scores that were defined as standards. In other words, different score points were found for each agency. However, the spread was not very large. In order to give an example of the kinds of fitness standards defined, the spread of median scores for each test of federal, state and municipal agencies is presented:\*

<b>Test</b>	<b>Range</b>
1.5 mile run	14:40 - 15:54
300 meter run	64.3 - 66.0
1RM bench raw score	151 - 165
1RM bench ratio	.78 - .84
Push up	25 - 34
Sit up	30 - 38
Vertical jump	15.5 - 16.0
Agility run	18.1 - 18.2
Sit and reach	15.0 - 16.7**

**\*These scores are provided as examples and can not be applied as standards by another agency unless a transportability validation is performed.**

\*\* The sit and reach was found predictive in less than 5% of the agencies

### **ISSUES REGARDING JOB TASK SIMULATION VERSUS FITNESS TESTS**

There are two distinct models with differing test modalities for validating physical performance tests. In general, public safety agencies use two types of tests: Job-task simulation tests (JTST) and physical fitness tests (PFT). Each type of test has advantages and disadvantages.

## **Job-Task Simulation Test**

For many, the biggest advantage to a JTST is that one can more easily recognize JTST as being job related. The test normally contains items that are specific job tasks a police officer may have to perform. Most versions of a JTST require the police officer to perform individual tasks such as dragging a victim to safety, making a pursuit etc. As long as the items are content valid (i.e. they contain frequent or critical tasks defined for the job) JTST can be defended for being job related.

JTST is not a new approach and agencies have been applying such tests for over 40 years. Historically, however many law enforcement applications of JTST have not been sustained with the popularity of JTST ebbing and flowing because of several drawbacks associated with JTST. Firefighter applications have tended to sustain the use of JTST because it is easier to quantify the physical demands for that class of employees as opposed to law enforcement officers.

- It is very difficult to measure or predict ability to perform all or even most physical job tasks. In fact, a typical JTST accounts for only 20-25% of the critical physical functions of law enforcement officers but does account for 50-75% of fire fighter essential functions.
- JTST are somewhat dependent on prior learning, and are therefore less predictive of overall fitness for duty. An applicant is being measured on a skill that will be learned in training. Consequently, it is often difficult to defend applicant standards based on a JTST.
- A JTST does not measure the underlying physical abilities or physical fitness. Consequently it does not provide a good overall measure of one's physical capabilities.
- Police officers tend not to train for JTST. Therefore, these tests do not necessarily encourage physical fitness or link directly with physical fitness programs.
- While you can't require public safety officers to be healthy, JTST do not have a bearing on reducing known health risks.
- JTST require a greater logistical effort to administer than fitness tests.
- More police officers suffer injuries using this type of testing as opposed to traditional physical fitness testing.
- Agencies have difficulty determining a passing score. The slowest incumbent time is often used or a statistical measure such as the 10%tile score. In some instances passing scores are just made up. None of those approaches appear to be defensible as being job related and consistent with business necessity.

## **Physical Fitness Tests**

PFTs biggest disadvantage is that it is somewhat more difficult to understand as being job related. The test items consist of accepted measures of physical fitness as underlying physical abilities. But since a police officer does not do push ups or sit ups as job tasks the validation of a PFT must be based on predictive or criterion validity. A PFT must be

validated through a construct/criterion validation processes to document the physical fitness levels (measured by the PFT) as being underlying and predictive of job performance of physical tasks. A criterion validation requires more effort, resources and cost.

PFT testing has been popular within law enforcement agencies and to some extent fire departments for the last 30 years because of several advantages to this approach.

- PFT discriminates well between those who can and cannot perform all the essential functions of the job.
- PFT predicts trainability in training.
- PFT links directly with a fitness program, encouraging fitness development.
- PFT has a direct bearing on lowering health risk and injuries.
- PFT tends to be easier to administer.
- PFT tend to account for up to 80% of the underlying physical abilities to perform most critical physical tasks.
- PFT can facilitate developing remedial fitness programs, as weak areas are immediately identifiable.

While FitForce and its associates prefer fitness testing for the reasons noted above, either type of test is defensible if validated as being job related and consistent with business necessity.

### **A Final Note**

Although we have stated and a recommended a preference for physical fitness testing, our recent experience with the fire science has challenged our position. Specifically, our recent recommendations have been to prioritize job task simulation testing for the fire service. This is due to the nature of the occupational demands of the fire service and the mandate to adopt the lowest level of fitness that predicts the ability to do the job.

#### **About FitForce™**

FitForce™ is committed to finding Total Fitness Solutions for the public safety community we serve. Our pledge is to provide our clients with the very best scientific, legal and practical training and education, validation of selection and retention standards, policy and procedure analysis and development, as well as ongoing administration, arbitration and litigation support.

FitForce™ and its consultant team together have over 60 years of public safety physical fitness experience. This includes: over 100 articles, columns, chapters, books and technical reports, experience with over 200 agencies and their representatives, over 130 validation studies and a database of over 4000 randomly selected law enforcement officers at agencies for whom we've developed standards. *If you would like to discuss how we could be of assistance to your agency, please call us at 978.745.3629.*