

SECTION G

TEST BATTERY AND SCORE STANDARDS DEFINITION

The various data analyses helped identify potential fitness test items but not criterion cutoff scores. Further analysis is required to 1) narrow the test battery to only those tests that are truly predictive of criterion performance on the job task simulation tests and 2) identify criterion cutoff scores for those fitness tests. From a criterion validity perspective, the judgment process must start with the identification of criterion cutoff scores for the job-task tests. The job-task scenario tests demonstrated content validity based upon the job analysis data and verification by the supervisor discussion group. Subject matter expert observation established a pass/fail cutoff score. As expected, the various data analyses clearly demonstrate that more fit officers score higher on the job-task tests. The challenge is to identify the cutoff point which differentiates between officers who can do the job and those who cannot. The job task simulation tests serve as criterion tests that define the ability to do the strenuous physical tasks of the job. Identification of the "criterion" cut point of acceptable job performance of physical tasks on those tests requires a structured process.

The ultimate selection of a standard cutoff must strike a balance among three elements:

1. What level of physical fitness is the minimum threshold to give reasonable assurance of safe and successful performance of frequent and critical job-related physical tasks?
2. What level of fitness is required to give reasonable assurance that a reserve of physical fitness is available for the most demanding critical tasks?
3. What level of performance is a fair and job-related expectation for all trainees and officers to achieve.

RATIONALE FOR STANDARDS DEFINITION

The rationale for the standards development process is as follows:

Standards should be based on statistics generated from an appropriate sample reflective of the population of officers. To be appropriate, the sample must be representative of the total class of employees.

Any standard identified should be predictive of officers' ability to perform essential job-tasks, especially those that were critical.

The impact of the standard on the incumbent population needs to be accounted for in the implementation of standards. It must be recognized that the majority of incumbent officers are performing adequately.

The cutoff score for any test should maximize predictability. The cutoff point that maximizes predictability is the one that most accurately classifies individuals based on their scores on the fitness tests. In other words, **a valid cutoff score is one in which those people who pass the fitness test also achieve the cutoff score for the criterion test and those who do not pass the fitness test do not pass the criterion test.** The terms applied to assess that predictability of a cutoff score are **specificity** and **sensitivity**.

The higher the specificity of a test score, the more it minimizes the possibility of having someone passing the fitness test but failing the criterion test. That type of person would be called a false positive. A test with good specificity helps insure that someone who passes the fitness test can perform the physical demands of the job. **It minimizes the risk of passing someone who cannot do the job.**

The higher the sensitivity of a test score, the more it minimizes the possibility that someone fails the fitness test but passes the criterion test. That type of person is called a false negative. A test with good sensitivity helps insure that someone who does not pass the fitness test is, in fact, someone who can not perform the physical demands of the job. **It minimizes the risk of failing a person that can do the job.**

The ideal test cutoff level would be one with 100% specificity and 100% sensitivity - that is, there would not be any false positives or false negatives. However, the reality of any type of testing is such that it is virtually impossible to achieve 100% specificity and sensitivity. Consequently, the judgment team had to decide which had the highest priority - specificity or sensitivity.

The judgment team, in evaluating the data, concluded that **specificity was the highest priority**. We concluded that the critical nature of an officer's mission was such that minimizing false positives was the priority. In other words, it is more important to have a test cutoff level that minimizes the risk of having someone pass the fitness test but fail in performing criterion job-tasks.

In conclusion, we based the rationale for setting fitness standards on having standards that insure officers can meet the physical performance demands of tactical situations. The various statistical analyses and comparisons are the methods for validly defining those standards.

The process for defining the cut points for expected officer performance involves two major steps: 1) defining the criterion test cutoff score and 2) defining the fitness test cutoff score.

DEFINING POTENTIAL CRITERION (JOB-TASK) TEST CUTOFF SCORES

Before defining a cutoff score for the fitness test, it is first necessary to define the cutoff score for the criterion test. We assumed that the faster an officer could perform the various tasks, the higher the probability he or she could successfully accomplish the mission.

The **first issue or question** is to establish the criterion score for each job- task scenario test (roadway clearance, extraction, pursuit and subdue). While the job task

scenario situations defined for the testing situation of this project have been used before, there are not any previously set cutoff levels. The judgment team defined two elements for determining levels against which to compare physical fitness test specificity and sensitivity.

Element # 1

The criterion score cutoff was based on the actual performance of the sample performing the job task scenario. Using the assumption that the majority of officers can perform a variety of job task situations in a satisfactory fashion, the 10th percentile level, 1 standard deviation below the mean (approximately the 16th%tile), and the 20th percentile level of performance on all tests were selected as three optional cutoff levels. In other words, the 10th%tile criterion assumes that 90% of the sample are performing adequately, the 16th%tile criterion assumes that 84% of the sample are performing adequately, and the 20th%tile criterion assumes that 80% of the sample are performing adequately. Those scoring at the 10th%tile took longer to complete a given job task scenario than those at the 20th%tile. In other words, the 100th%tile would reflect the best performance (fastest time) and the 1st%tile would reflect the poorest performance (the slowest time).

Additional rationales for the selection of these cutpoints are the conventional practices in the field and conclusions from past validation studies. First, there is a consensus assumption within the field that the faster the time to perform strenuous physical tasks, the more effective the performance of the task. Since the tasks utilized are %critical+tasks where injury or loss of life could be potential consequences, that rationale has been accepted both by professionals in the field and the court.

The standard deviation (sd) is a statistic that reflects the variation of test scores around the average score. It is generally accepted as a major cutpoint for viewing the significant differences between scores. A standard deviation cutpoint is often used as an indicator of acceptable performance. Likewise, past validation studies conducted by others in the field as well as ourselves, have found that either the 1sd, 10th or 20th percentile scores consistently appear as the indicators of minimum or acceptable performance.

Element # 2

During the validation testing, the trained Coordinators who administered the tests (subject matter experts) was asked to observe subjects undergoing each job task scenario test and make a judgment as to whether the performance of each subject would be considered effective or not effective for completing the scenario. The supervisors were trained to use a rating process of effective or non-effective

performance based on the subject's ability to 1) perform the job tasks with the appropriate skill level and procedures, 2) perform the job tasks in a safe and efficient manner and 3) perform the job tasks at a pace required to accomplish the mission of the scenario successfully. In turn, they were encouraged to discuss their ratings to come to a consensus rating when possible. An example for the pursuit scenario is given below:

SCENARIO # 3

RATING GUIDELINES Use these questions to help decide on effectiveness/non effectiveness

Rate the individual as ineffective if the following performance is noted:

- * The time it takes is too long
- * Can not make it over, under or through an obstacle
- * Walks too much through the course or up and down the stairs
- * Can not perform the restraining tasks proficiently (cant deliver forceful blows)
- * Uses poor technique which would limit the ability to perform the tasks in a real life situation

_____ Effective

_____ Ineffective

For the officers tested, the times for those rated as ineffective were compared to the times for those rated effective. The range of ineffective ratings for the first and second scenarios were so large and variable that it was impossible to draw a relationship to the times to perform those scenarios. However, on the third scenario the ineffective rating times were slower and tended to cluster between the 10th and 16th %tile. This provided concurrent validity with past studies indicating that is the usual range of ineffective performance. Consequently we used a range of percentile performance ranks to define criterion performance.

Utilizing all the potential cut points yielded the following cutoff criterion scores presented in Table G1.

TABLE G1
POTENTIAL CRITERION CUTOFF LEVELS FOR THE JOB TASK SCENARIO TESTS

	<u>Roadway clearance</u>	<u>Extraction</u>	<u>Pursuit and subdue</u>	<u>Total time</u>
20th %tile	39.4 sec.	25.9 sec.	4: 05	5:13
1 sd.(16th %tile)	40.6 sec.	27.2 sec.	4: 20	5:23
10th %tile	42.9 sec.	35.4 sec.	4: 30	5:34
Ineffective level	na	na	4: 25	na

We considered each of those scores as a criterion cutoff for performing specificity and sensitivity analysis. **Evaluating four criterion cutoff options provides a more meaningful view of the ability of the fitness tests to predict job performance.**

IDENTIFYING POTENTIAL FITNESS TEST CUTOFF SCORES

To identify potential physical fitness test cutoff scores, we applied the same rationale utilized in identifying criterion test (i.e., job-task scenario) cutoff levels. We considered the 10th, 1sd (16th), 20th, 30th, 40th, and 50th percentile of the officers scores on the selected fitness tests for specificity/sensitivity analysis. We selected the 10th to 50th tile because our previous experiences indicate that the most predictive scores will fall within that range. Table G2 shows those six performance levels for each test.

TABLE G2

	FITNESS TEST RAW SCORES							
	1.5mile <u>Run</u>	300-m <u>Run</u>	Push- <u>Up</u>	Sit- <u>Up</u>	Vert. <u>Jump</u>	1RMBP <u>Raw</u>	<u>Ratio</u>	Agility <u>Run</u>
50th	15:05	64.3 sec.	30	38	17.4in.	175lbs.	.88	17.9 sec.
40th	15:54	67.0 sec.	25	35	16.0in.	162lbs.	.82	18.3 sec.
30th	16:34	70.0 sec.	22	33	15.5in.	150lbs.	.75	18.9 sec.
20th	17:40	75.8 sec.	20	30	14.0in.	135lbs.	.67	19.5 sec.
16th	18:11	78.0 sec.	19	29	13.0in.	135lbs.	.63	20.0 sec.
10th	18:54	83.0 sec.	15	26	12.1in.	106lbs.	.60	19.8 sec.

SPECIFICITY AND SENSITIVITY ANALYSIS

Fitness Intervention Technologies performed sensitivity and specificity analysis by assessing the percentage of the testing sample correctly identified as passing and failing each of the criterion test cutoff point options for each fitness test cutoff point option. A way of viewing that statistical analysis is with the matrix that follows:

The condition is having the fitness and ability to meet/pass the job criterion.

Positive test = Passing the test indicating having the condition (fitness to do the job)

Negative test = Failing the test indicating not having the condition (fitness to do the job)

Sensitivity= The %tage of individuals with the **condition (meets/passes job criterion)** that are correctly identified by the test-i.e. pass the test as **having the condition (minimal level of fitness** to meet/pass the job performance criterion)

The higher the sensitivity the better at controlling **false negatives**- individuals who have the condition (meet/pass job criterion) but the test identifies them as not having the condition (minimal levels of fitness to meet/pass the job criterion). A test with good sensitivity helps insure that someone who does not pass the fitness test is, in fact, someone who can not perform the physical demands of the job. **It minimizes the risk of failing a person that can do the job.**

Low sensitivity will mean that the test may fail individuals who if fact have the minimal

levels of fitness to meet/pass the job criterion.

Specificity= The %tage of individuals **without the condition (meets/passes job criterion)** that are correctly identified by the test-i.e fails the test as **not having the condition** (minimal levels of fitness to meet/pass the job criterion).

The higher the specificity the better at controlling **false positives**- individuals who do **not** have the condition (meet/pass job criterion) but the test identifies them as having the condition (minimal level of fitness to meet/pass the job criterion). A test with good specificity helps insure that someone who passes the fitness test can perform the physical demands of the job. **It minimizes the risk of passing someone who cannot do the job.**

Low specificity will mean that the test may pass some individuals who may in fact do not have the minimal levels of fitness to meet/pass the job criterion.

Fitness Test	Criterion Test	
	(Has condition) Passes Criterion (+)	(Does not have condition) Fails Criterion (-)
Pass Fitness (+)	A = True Positive (Pass both tests)	B = False Positive (Pass fitness but fail criterion test)
Fail Fitness (-)	C = False Negative (Fail fitness but pass criterion)	D = True Negative (Fail both tests)
Sensitivity =	A / A+C	D / B+D
Specificity =		

The scores for specificity and sensitivity reflect the percentage of individuals correctly identified as passing and failing the criterion test. The higher the percentage for cells A and D, the greater the predictability for minimizing the risk of misclassifying an individual. The specificity and sensitivity analysis produces percentages of accuracy for each category:

The **specificity** percentage reflects how accurately a given fitness test cut point predicts who will also fail the criterion test. For example, a specificity

percentage of 80% means that 80% of those who failed the criterion test also failed the fitness test at its cut point.

The **sensitivity** percentage reflects how accurately a given fitness test cut point predicts who will also pass the criterion test. For example, a sensitivity percentage of 70% means that 70% of those who passed the criterion test (at that cutpoint) also passed the fitness test at its cut point.

A formal step-by-step analysis of the specificity and sensitivity data defines fitness tests cut points. To be considered as a potential cut point or standard, each fitness level had to have both a specificity and sensitivity of at least 70%. In other words, as a minimum, the cut point provides 70% accuracy of predicting passing and failing.

In conclusion, the specificity and sensitivity analysis is the statistical method to determine a cut point standard with the most validity. It is the method to determine the score that maximizes the assurance of an officer's capability to perform physical tasks and maximizes the fairness of the standard.

Fitness scores

The specificity and sensitivity analyses required completing a statistical procedure for multiple combinations. Each job-task scenario required the following computations: The selected cut points for each fitness test (10th, 16th, 20th, 30th, 40th, and 50th %tiles) were analyzed for specificity and sensitivity for the job-task cut point (effectiveness level, 10th%tile, 16th%tile, 20th%tile) for each scenario plus total score.

Tables G3, G4 and G5 present the specificity and sensitivity percentages for scenario # 2, #3 and the total score. Scenario # 1 did not have any specificity/sensitivity percentages that met the 70/70 criterion. Each table shows the

combinations of fitness test and potential job-task cut points whose specificity and sensitivity percentages met the 70% criterion.

TABLE G3
SPECIFICITY AND SENSITIVITY FOR FITNESS TESTS RAW SCORES
AND JOB TASK SCENARIO 2 CRITERION CUTOFF SCORES

Scenario 2 Criterion Cutoff	Fitness Test Cutoff	1.5 mile		Push		Sit		Vert.		1RBP		1RBP		Agility
		<u>Mile</u>		<u>Up</u>		<u>Up</u>		<u>Raw</u>		<u>Ratio</u>		<u>Run</u>		
		Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp	
10%tile 10%														
10%tile 16%														
10%tile 20%														
10%tile 30%						76	72							
10%tile 40%						74	83							
10%tile 50%														
16%tile 10%														
16%tile 16%														
16%tile 20%														
16%tile 30%														
16%tile 40%														
16%tile 50%														
20%tile 10%														
20%tile 16%														
20%tile 20%														
20%tile 30%														
20%tile 40%														
20%tile 50%														

TABLE G4
SPECIFICITY AND SENSITIVITY FOR FITNESS TESTS RAW SCORES
AND JOB TASK SCENARIO 3 CRITERION CUTOFF SCORES

Scenario 3 Criterion Cutoff	Fitness Test Cutoff	1.5 mile		Push		Sit		Vert.		1RBP		1RBP		Agility
		<u>Mile</u>		<u>Up</u>		<u>Up</u>		<u>Raw</u>		<u>Ratio</u>		<u>Run</u>		
		Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp	

Effective	10%			
Effective	16%			
Effective	20%			
Effective	30%	78 95		
Effective	40%		70 80	
Effective	50%			

10%tile	10%			
10%tile	16%	91 77		
10%tile	20%	85 78		
10%tile	30%	77 94		
10%tile	40%			
10%tile	50%			

16%tile	10%			
16%tile	16%			
16%tile	20%	88 71		
16%tile	30%	81 92		
16%tile	40%	70 96	73 78	
16%tile	50%			

20%tile	10%			
20%tile	16%			
20%tile	20%			
20%tile	30%	84 88		
20%tile	40%	73 94	76 80	72 72
20%tile	50%			

TABLE G5
SPECIFICITY AND SENSITIVITY FOR FITNESS TESTS RAW SCORES
AND TOTAL CRITERION CUTOFF SCORES

Scenario Criterion	Fitness Test	1.5 mile <u>Mile</u>	300-M <u>Up</u>	Push <u>Up</u>	Sit <u>Jump</u>	Vert. <u>Raw</u>	1RBP <u>Ratio</u>	1RBP <u>Run</u>	Agility
-----------------------	-----------------	-------------------------	--------------------	-------------------	--------------------	---------------------	----------------------	--------------------	---------

<u>Cutoff</u>	<u>Cutoff</u>	Se	Sp														
10%tile 10%																	
10%tile 16%																	
10%tile 20%		86	83														
10%tile 30%		77	94					76	72			75	72				
10%tile 40%				70	83			74	74								
10%tile 50%																	
16%tile 10%																	
16%tile 16%																	
16%tile 20%																	
16%tile 30%		80	89														
16%tile 40%	70	96		73	82			76	71								
16%tile 50%																	
20%tile 10%																	
20%tile 16%																	
20%tile 20%																	
20%tile 30%		84	86														
20%tile 40%	72	91		77	83												
20%tile 50%																	

JUDGMENT PROCESS FOR SELECTING TEST CUT POINTS

A formal step-by-step process was implemented to review the specificity and sensitivity data to define fitness tests cut points. There were two (2) steps:

- Categorization of potential fitness test cut points
- Definition of fitness test scores for each cut point

1. Categorization of potential cut points. We first looked at potential cutpoints that had high specificity and sensitivity. Because a judgment process must be applied it is difficult to assure a definitive criterion cutpoint for the job task scenarios. Consequently, a band or range of criterion cutpoints (effective rating, 10th%tile, 16th%tile, 20th%tile) was employed. The fitness test percentile that had the highest specificity (with a minimum of 70% sensitivity) for each of the three potential criterion cutpoints across each of the three job task scenarios (roadway clearance, extraction, and pursuit and subdue and total score) was averaged for each fitness test and rounded off to the nearest decile. The same process was applied for the fitness test percentile that had the highest sensitivity (with a minimum of 70% specificity). These provide reasonable cutpoints based

on all the data. These are presented in Table G6.

TABLE G6
FITNESS SCORE RAW SCORE PERCENTILE FOR SPECIFICITY CRITERION CUT POINTS

Fitness test	1.5 Run	Push Up	Sit Up	Vert. Jump	1RMB Ratio
Specificity cutpoint %tile		40th	40th	40th	40th 30th
Sensitivity cutpoint %tile		30th	40th	40th	30th 30th

2. Definition of fitness test score for each cut point. As can be seen the specificity and sensitivity cutpoints for each test are identical except for the 1.5 mile run and vertical jump. The various raw score cut points are presented in Table G7 that correspond to the percentiles determined to predict job performance shown in Table G7.

TABLE G7
FITNESS TEST RAW SCORE CUTPOINTS ACROSS CRITERION PERCENTAGE CUTPOINTS

1.5-mile run	(40th%tile)	= 15:54
Push up	(40th%tile)	= 25
Sit up	(40th%tile)	= 35
Vertical jump	(40th%tile)	= 16 inches
1RM bench ratio	(30th%tile)	= .75 (weight pushed is 75% of body weight)

RECOMMENDED PHYSICAL FITNESS TEST BATTERY

The previous Section defined a potential fitness test battery based upon the job task analysis and relationship data. That battery consisted of items that demonstrated criterion validity in that those test items were predictive of performance on the job task simulation tests at a general level i.e. the higher the fitness test score the better the performance on the job task simulation test. Those test items were as follows:

- 1.5 mile run
- 300 meter run

- Maximum push ups
- 1 minute sit ups
- Vertical jump
- Illinois agility run
- 1 RM bench press raw and ratio score

Those test items only demonstrated predictability, at a imprecise level, to performance on the job task simulation tests. The specificity and sensitivity analysis provides data to determine which tests have cutpoints that are reasonably predictive of the specific "criterion" cutpoints on the job task simulation tests. The recommendation is that only those test items that have practicability of the specific criterion cutpoints on the job task (criterion) simulation test should be part of the fitness battery. With that assertion, the recommended test battery should consist of the following items:

- 1.5 mile run
- Maximum push up
- 1 minute sit up
- Vertical jump
- 1 RM bench press ratio

The same test battery should be applied for applicants and incumbents. However, the incumbent test battery should also contain the sit and reach test and the body fat test. While not areas for job related fitness standards they do represent important health related fitness areas that officers should be given feedback on.

RECOMMENDED JOB RELATED PHYSICAL FITNESS STANDARDS

The objective of this validation study was to define a job related fitness battery and job related fitness standards that provide maximum predictability for insuring an individual can perform the strenuous physical demands of the job. Likewise, a

secondary objective was to define tests and standards that could provide maximum defensibility for job relatedness. To meet those objectives the studiesqdata direct the definition for recommended tests and standards. The recommended standards of performance for the fitness test battery should be those test cutpoints that provide maximum predictability. Those standards were:

1.5-mile run	= 15:54
Maximum push up	= 25
1 minute sit up	= 35
Vertical jump	= 16 inches
1RM bench ratio	= .75

STANDARDS COMPARISON TO OTHER NORM STANDARDS

From a comparative perspective the recommended fitness test cutpoints can be contrasted against other fitness norms that have been applied as standards for incumbent law enforcement officers. This can add some meaning to the potential difficulty level of the cutpoints (as standards). Comparisons are provided in Table G8 with the FitForce and Cooper Institute for Aerobics Research (CIAR) single standard police norms, CIAR single standard general population norms, and CIAR age and gender based general population norms. The percentile rank at which the standard would fall on in each of the norm profiles is recorded.

TABLE G8
STANDARD COMPARISONS TO OTHER FITNESS TEST NORMS

TEST	CIAR Police norms	CIAR Single norms	CIAR/FitForce age and gender population standards							
			male				female			
			20-29	30-39	40-49	50 +	20-29	30-39	40-49	50 +
1.5	25th	25th	5th	5th	15th	35th	30th	40th	60th	80th
Push up	30th	65th	25th	45th	65th	80th	70th	85th	99th	na

Vert.j.	30th	na	10th	15th	45th	70th	65th	90th	99th	na
BP ratio	10th	55th	5th	15th	25th	50th	70th	85th	90th	99th
Sit up	65th	70th	30th	45th	65th	80th	50th	80th	90th	99th

As can be seen, the standards fall below the average (50th%tile) on the single standard police norms, except for the sit up test, and higher on the general population norms except for the 1.5 mile run. However, there is a different trend when viewing the CIAR age and gender based norms. Basically, the standards are well below the 50th%tile for the younger males and above the 50th%tile for older males (above age 50) for the vertical, push up and sit up and at the 50th %tile or below for the 1.5 mile run and 1RM bench press ratio. For females, the standards are at the 50th %tile or above for all ages on all tests except the 1.5 mile run. 1.5 mile standard is below the average female performance at the 20-39 age range.

These trends would indicate that the recommended standards would demonstrate adverse impact against females if that gender classification were viewed as a random sample of the population. However, from a training perspective these standards are obtainable if one were to train regardless of age or gender.

One additional comparison can be made against prior validation studies whereby standard cutpoints for the fitness tests that have been validated. Table G 9 compares the recommended standards with those validated from studies with municipal law enforcement agencies.

TABLE G 9
STANDARD COMPARISONS TO OTHER FITNESS STANDARDS

<u>Test</u>	<u>URMMA data</u>	<u>Previous studies</u>
-------------	-------------------	-------------------------

1.5 mile run	15:54	14:26 - 16:38
Push up	25	20 - 25
Sit up	35	18 - 35
Vert.jump	16 in.	13.0 - 18.5 in.
1RM bench ratio	.75	.64 - .82

This study's recommended standards tend to fall at the midpoint for 1.5 mile run, vertical jump and 1RM bench press ratio. The push up and sit up standards tend to be at the high end of the range.

It must be recognized that these standards represent low levels of fitness when viewed strictly as fitness measures. However, these trends must be viewed in the context of job relatedness in that **the standards are established to only guarantee minimal fitness to perform the strenuous tasks of the job.**

CONCLUSION

The recommended fitness battery and standards should be applied as both an entrance and incumbent requirement. Incumbents should also receive feedback on two additional tests. As mentioned previously, the body fat test is a valuable tool for providing officers feedback on an important area, even though it does not add much predictive power to the cardiovascular and muscular strength predictors. As a consequence, we recommend it be included in the trainee battery as an educational tool, but not as a standard. We recommend a voluntary goal of the 50th%tile on the Cooper age and gender based norms. Likewise, we recommend that the sit and reach test (flexibility) also be included as a voluntary goal measurement area for officers. The rationale is that a major justification for a fitness program is to reduce health risk. The sit and reach test measures an important fitness area: lower back and hamstring flexibility. We recommend a voluntary goal of 50th%tile on the Cooper age and gender

norms. Those norm based standards are as follows:

TEST	MALE				FEMALE			
	20-29	30-39	40-49	50 +	20-29	30-39	40-49	50 +
Body fat (%)	19.0	19.0	21.1	22.7	22.1	23.1	26.4	30.1
Sit and reach (in.)	17.5	16.5	15.5	14.5	20	19	18	17.5

STANDARDS APPLICATION CONCERNS

By having single job-related standards for all incumbent officers and applicants, the 19 agencies should prevail if challenged under Title VII of the 1964 Civil Rights Act, under the Civil Rights Act of 1991, under the ADA, or under the ADEA. The standards are based on a sample of existing officers. As such, a standard corresponding to the 40th %tile of officer performance (as an example), means (by definition), that 40% of existing incumbents or recruits would fail that standard at present. However, at a general level, it is reasonable to assume that with training, presently unfit recruits and incumbents could reach the standards in a reasonable time frame.

Since these job-related standards will be applied to applicants, recruits, and incumbents, careful application and ongoing review is appropriate. The standards recommended meet the priorities of having job related requirements that predict physical performance capabilities and are defensible if challenged. We have attempted to follow the Uniform Guidelines in implementing the validation process and in defining those standards. However, we recognize that the agencies has many other concerns and objectives that must be addressed other than those we have expressed as priorities. Consequently, the recommended standards can be viewed as **eventual standards** to apply after the other concerns and priorities are addressed by the . These will be covered in the Implementation Section I.

EXTRA STUFF

The rationale for determining the fitness test cutpoint that demonstrated to highest specificity with acceptable sensitivity across all the scenarios is as follows:

OPTION # 1 Since officers are expected to perform all, including the most demanding tasks the fitness test cutpoint should be at the highest specificity level demonstrated by the data regardless of the job task scenario. In other words. the cutpoint should reflect the highest demand. If a officer can meet that requirement then it can be concluded that the officer could meet all the other job task demands.

OPTION # 2 Since we do not have the data to assure a definitive criterion cutpoint for the job task scenarios, a band or range of criterion cutpoints (effective rating, 10th%tile, 16th%tile, 20th%tile) is employed. The fitness test percentile that had the highest specificity for each of the four potential criterion cutpoints across each of the three job

task scenarios (pursuit and subdue, extraction, and total time) was averaged for each fitness test. This provides a reasonable cutpoint based on all the data.

OPTION # 3 Since we do not have the data to assure a definitive criterion cutpoint for the job task scenarios, a band or range of criterion cutpoints (effective rating, 10th%tile, 16th%tile, 20th%tile) is employed. All of the fitness test percentiles that demonstrated specificity (with a minimum 70% sensitivity) for each of the four potential criterion cutpoints across each of the three job task scenarios (pursuit and subdue, extraction, and total time) was averaged. This provides a reasonable cutpoint based on all the data.