

# **Women, Combat, and Statistics**

*Linking combat performance & fitness*



Capt Kim Hale  
Maj Katie Batterton

AETC Studies and Analysis Squadron

# January 23, 2013

**Army:** 19 specialties including Infantry, Armor, Cavalry, & Special Forces

**Marine Corps:** 22 specialties including Artillery, Tank Officer, & Rifleman

**Navy:** 5 specialties including SEAL & Special Warfare

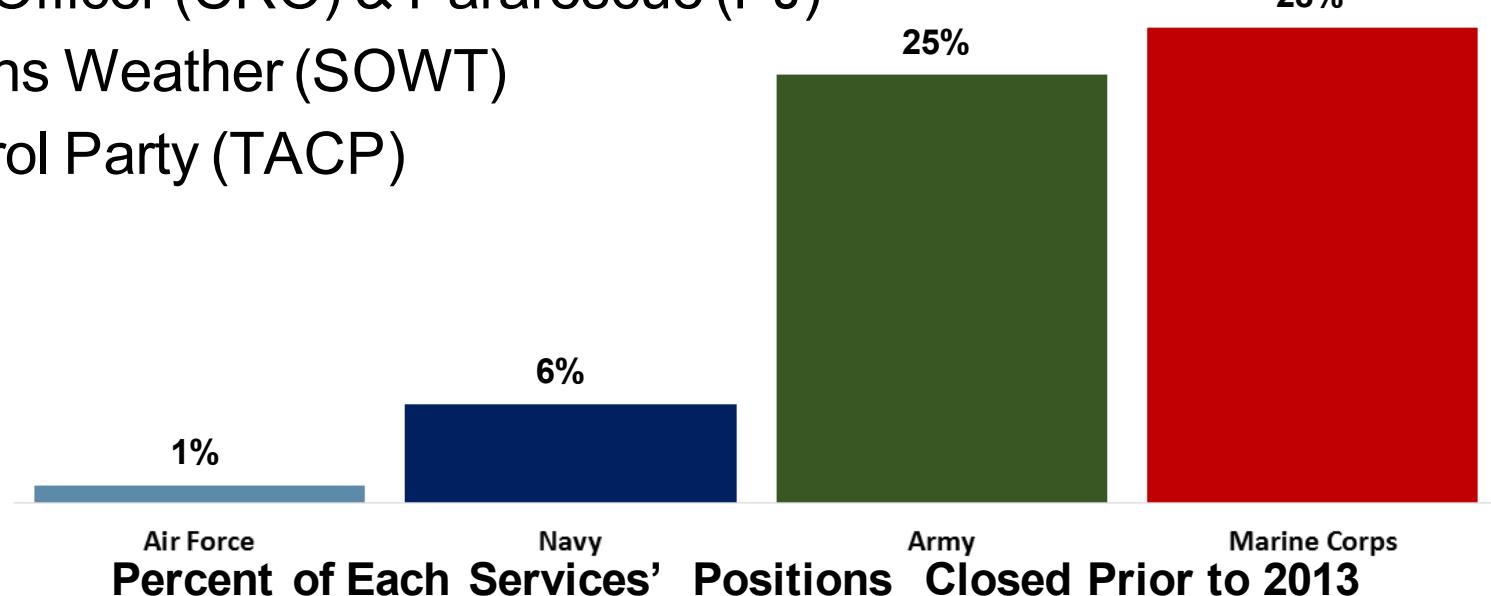
**Air Force: 6 specialties known as the “Battlefield Airmen”**

Special Tactics Officer (STO) & Combat Control Team (CCT)

Combat Rescue Officer (CRO) & Pararescue (PJ)

Special Operations Weather (SOWT)

Tactical Air Control Party (TACP)



# December 3, 2015

“As long as they qualify and meet the standards, women will now be able to contribute to our mission in ways they could not before. They'll be allowed to drive tanks, fire mortars, and lead infantry soldiers into combat.”

- *Secretary of Defense Ash Carter*

## *How did we get here?*

# Women in Service Review

Services directed to review standards & policies  
for each affected specialty

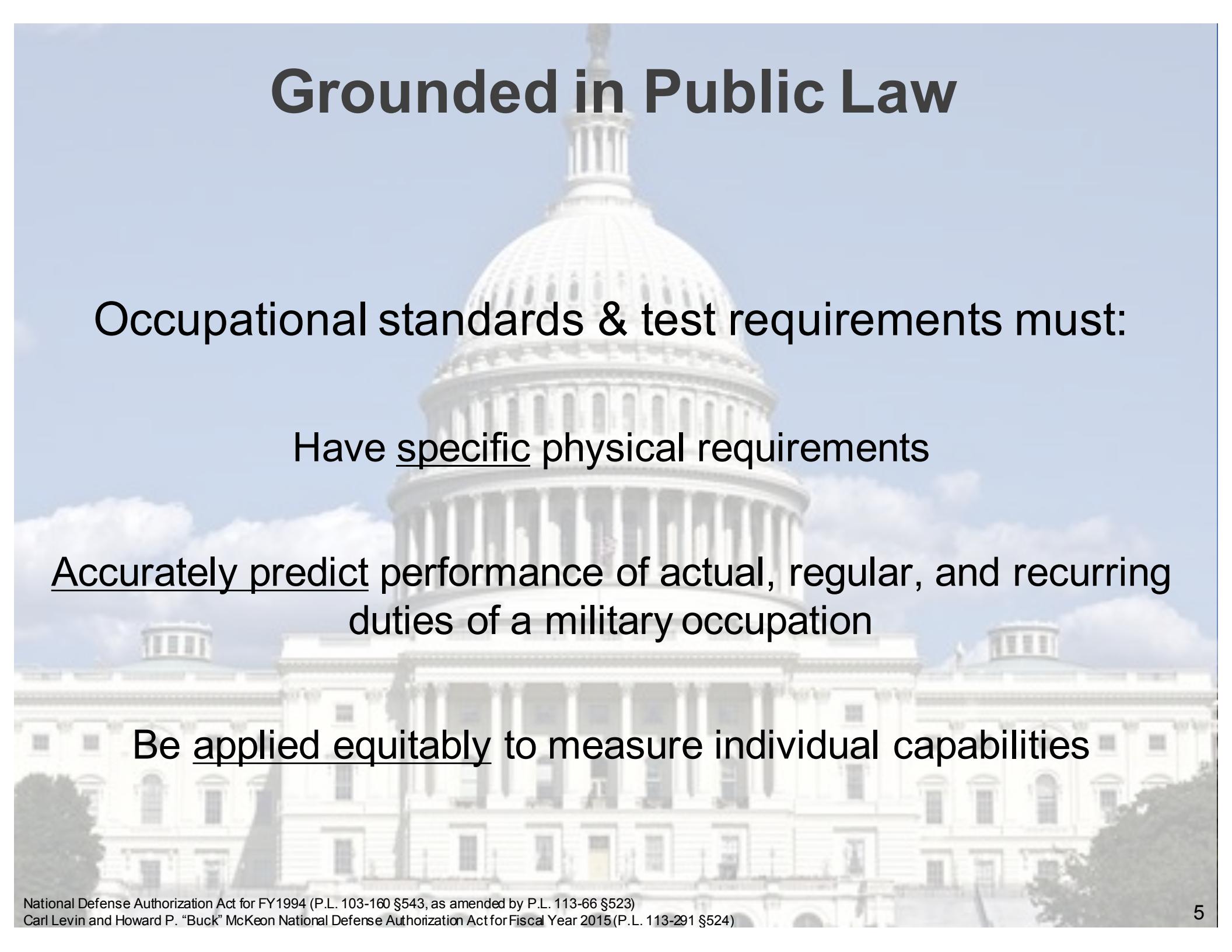
Validate performance standards

Standards must be gender-neutral

Training should reflect necessary  
knowledge, skills, and abilities



# Grounded in Public Law



Occupational standards & test requirements must:

Have specific physical requirements

Accurately predict performance of actual, regular, and recurring duties of a military occupation

Be applied equitably to measure individual capabilities

# Quantifying Combat Fitness

## Occupational Requirements Analysis

Used extensively in Canadian & British: Militaries, Police & Fire Departments

---

Define Critical Physical Tasks (CPTs)

Develop Physical Task Simulations (PTSs) from CPTs

Develop list of Physical Fitness Tests (PFTs) to consider

    Data collection and establish PTS – PFT link

    Determine standards to predict operator performance

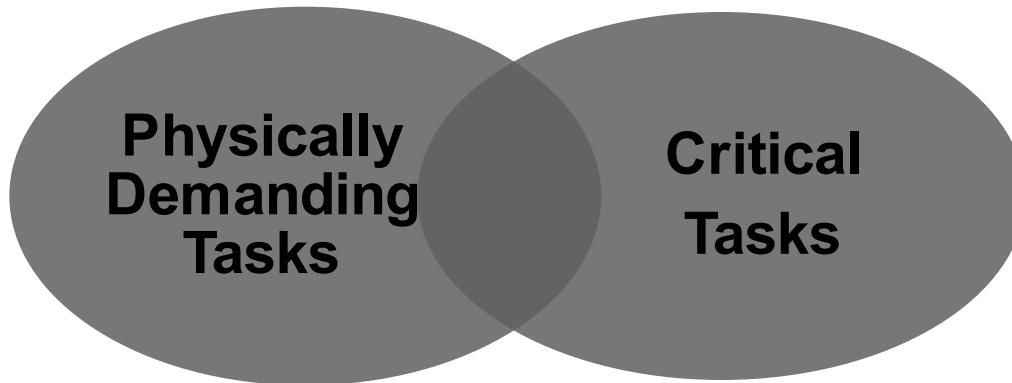


# Defining Critical Physical Tasks

Initial list of all work tasks from occupational analysis

Solicited BA input in focus groups, survey, and interviews

Identified critical and physically demanding tasks



## AFSC EXAMPLE TASKS

CCT-STO Execute down vehicle drills/Assist teammate from vehicle

PJ-CRO Recover/Remove survivors from water, wreckage, or vehicle

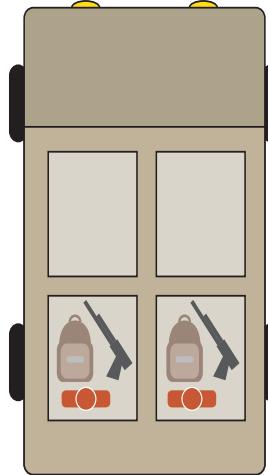
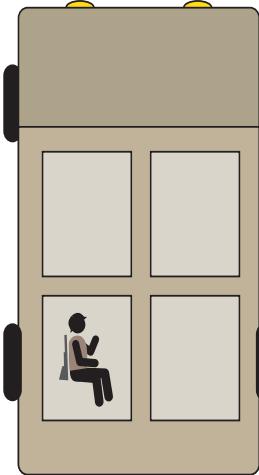
TACP Unload/Cross load personnel & equipment from disabled vehicle

# Developing Physical Task Simulations

*Example: Transfer Personnel and Equipment*

## Task

Move injured personnel/equipment



## Primary Measures

Completion (Y/N)

Time

Distance

## Secondary Measures

Heart Rate

Rated Perceived Exertion



# Developing Physical Fitness Tests

14 Physical Descriptors  
(e.g., *velocity, lift*)

+

6 Fitness Components  
(e.g., *muscle strength, anaerobic capacity*)

=

85 Physical Fitness Tests  
(e.g., *Farmer's Carry*)



Primary Measures  
Time/Reps/Distance/lbs

Secondary Measures  
Heart Rate

# Data Collection

April – July 2015

## Pilot Study (65 PFTs)

21 subjects

11 BA operators

## Main Data (56 PFTs)

171 subjects

56 BA operators

62 females, 109 males

## Validation Data (23 PFTs)

34 subjects

18 BA operators

9 females, 25 males

## Week 1: Physical Fitness Tests



## Week 2: Physical Task Simulations



# Reducing the PFT Items

*While Maximizing PTS Breadth and Minimizing PFT Redundancy*

Pearson correlation or Area Under the Curve

Data-based multiple regression models

- Limited to best 15 PFT items per PTS
- Linear or logistic

Physiological hypothesized multiple regression models

- Restricted to 5 PFT items per PTS
- AICc used as model selection criteria

*Statistical results were combined with other qualitative metrics  
for the final test battery selection.*

# Prototype Operator Test Battery



10 Physical Fitness Test Items

Covers all 6 physical fitness components

Covers 11 of 14 physical descriptors

---

Includes tests:

Familiar to AF - 1.5 mile run, pull-up

New to AF testing - weighted metronome lunges, 1000m row ergometer

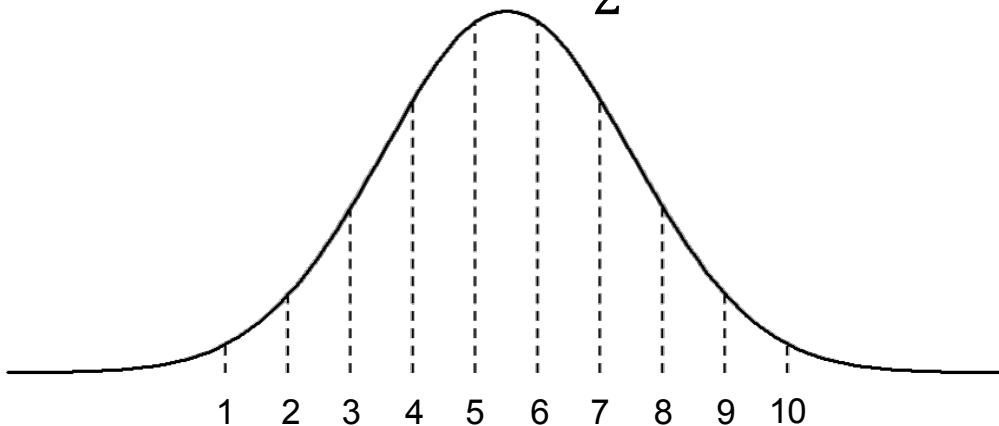
# Common Method – STEN Scales

For each test battery (TB) item:

Let  $i$  denote the point level,  $i = \{1, \dots, 9, 10\}$

then

$$\hat{x}_{i,TB} = (i - 5.5) \times \frac{S_{TB}}{2} + \bar{x}_{TB}$$



*Normative scale –  
score based on relative performance within population*

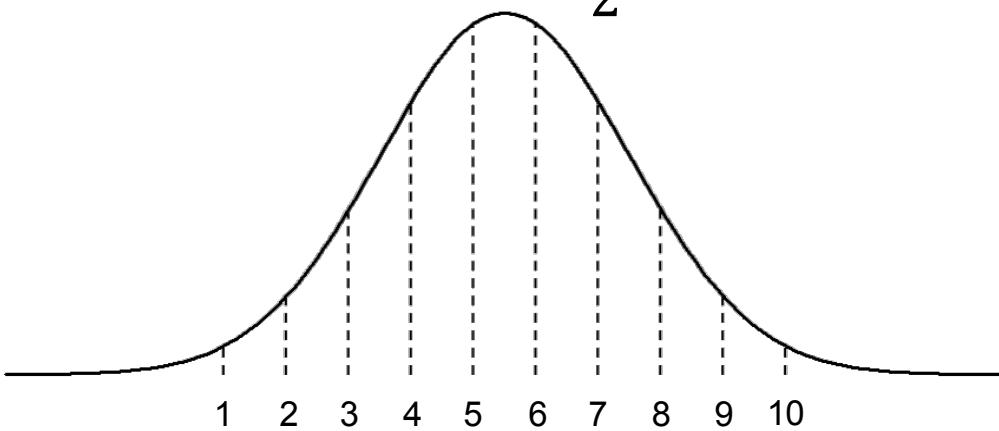
# Common Method – STEN Scales

For each test battery (TB) item:

Let  $i$  denote the point level,  $i = \{1, \dots, 9, 10\}$

then

$$\hat{x}_{i,TB} = (i - 5.5) \times \frac{S_{TB}}{2} + \bar{x}_{TB}$$



*Are we interested in normative behavior?*

*Does this scale meet the intent of public law?*

*Can we do better?*

**Normative scale –  
score based on relative performance within population**

# New Method – Anchored Scales

Univariate probit model links PTS completion and TB item performance

For each test battery (TB) item:

Let  $i$  denote the point level,  $i = \{1 \dots, 9, 10\}$

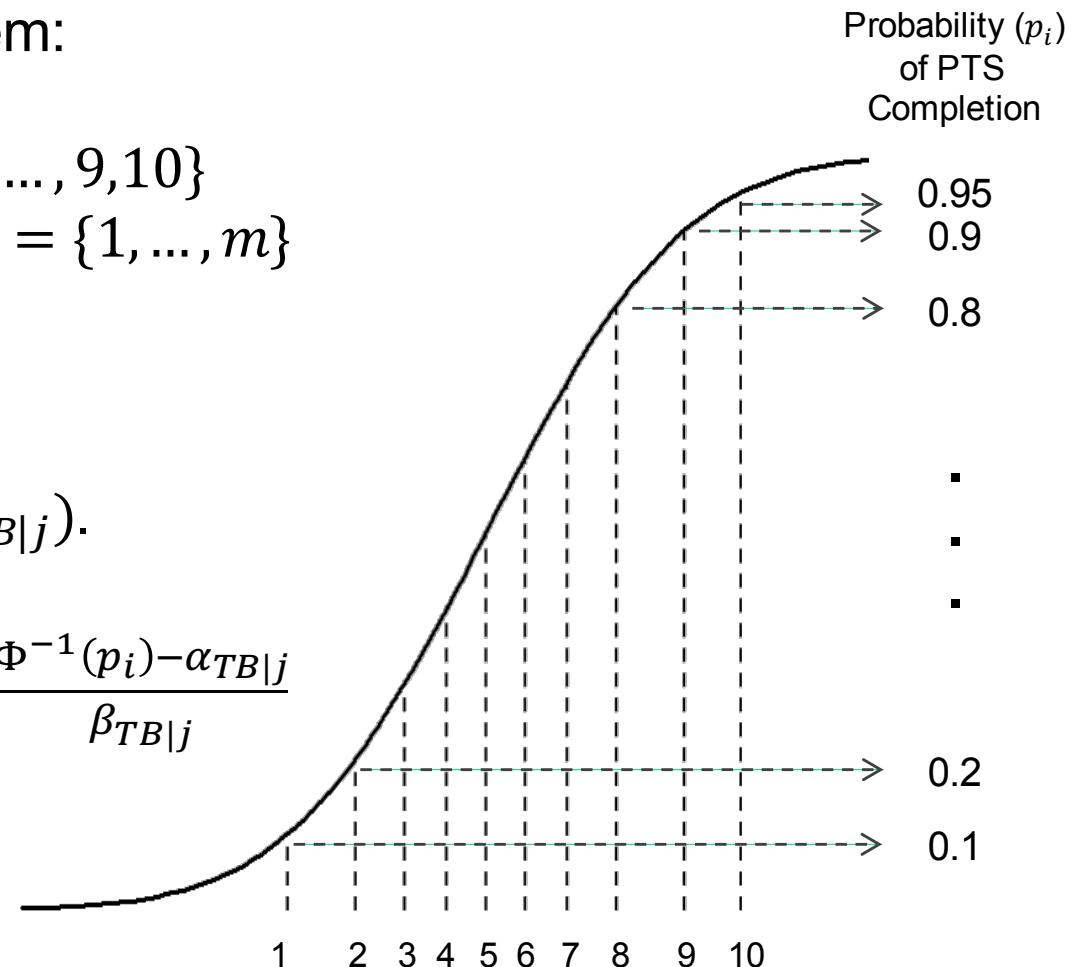
Let  $j$  denote a PTS linked to the TB,  $j = \{1, \dots, m\}$

then

$$\pi(x_{TB|j}) = \Phi(\alpha_{TB|j} + \beta_{TB|j}x_{TB|j}).$$

$$\text{now fix } \pi(x_{TB|j}) = p_i, \text{ then } \hat{x}_{i,TB|j} = \frac{\Phi^{-1}(p_i) - \alpha_{TB|j}}{\beta_{TB|j}}$$

$$\text{and } \hat{x}_{i,TB} = \sum_{j=1}^m \frac{\hat{x}_{i,TB|j}}{m}$$



*Criterion based scale –  
score linked to probability of success on task*

# Applying New Method

Repeated derivation of PTS - TB linkage for 10-item test

Reviewed extremes for physiological realism

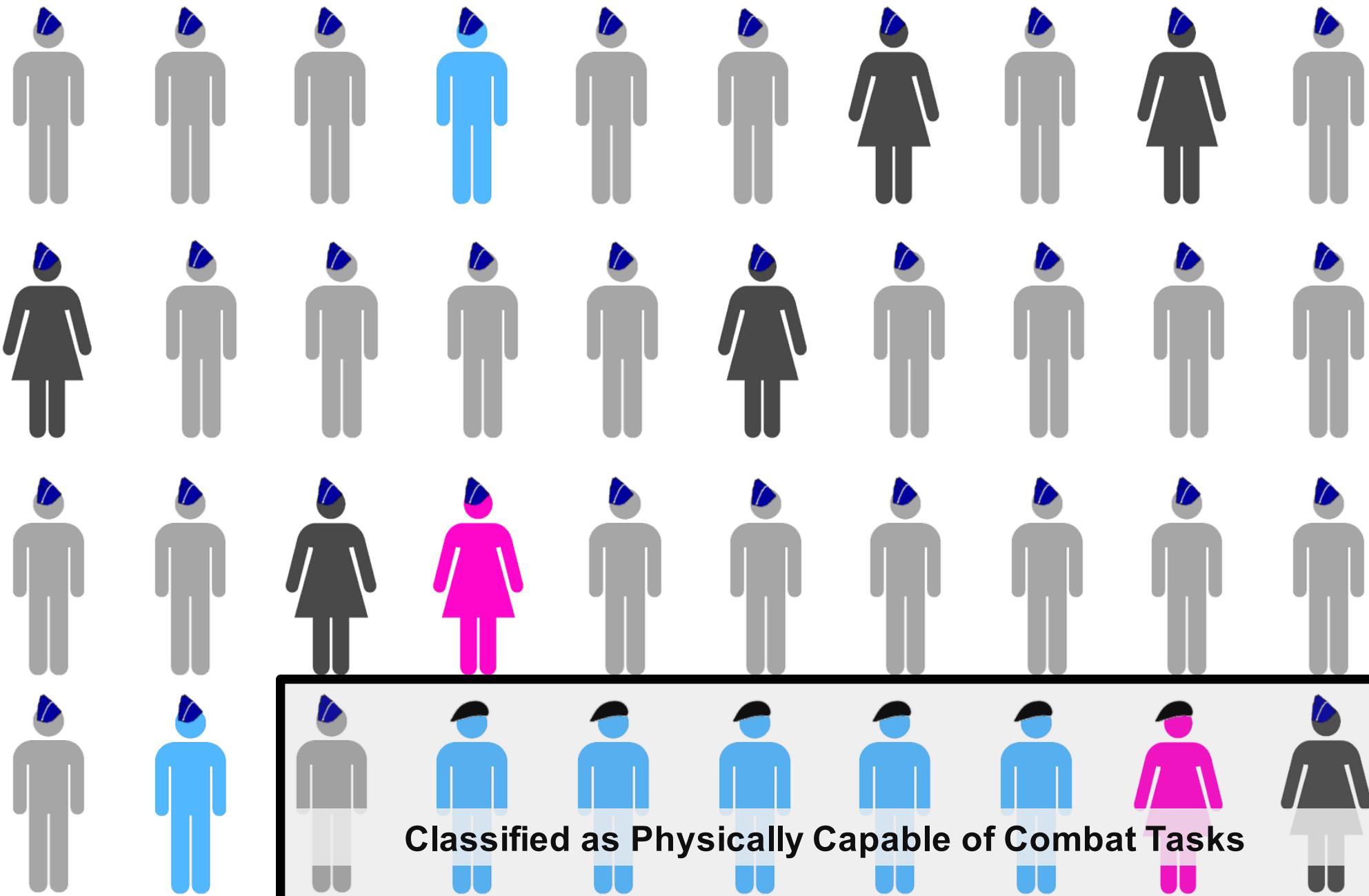
---

PTSSs Linked to Med Ball Toss Total	TB item: Medball Toss (Side/Back Total)		Operationally Relevant ↓ Anchored		
	PJ	TACP	STEN	PJ	TACP
Airfield Ops			9	51.1	46.5
Transfer Personnel & Equipment	Transfer Personnel & Equipment		5	30.6	37.5
Casualty Movement	Casualty Movement		3	20.4	39.0
Rope Ladder	Rope Ladder			34.0	35.5
Small Unit Tactics	Small Unit Tactics				

Points

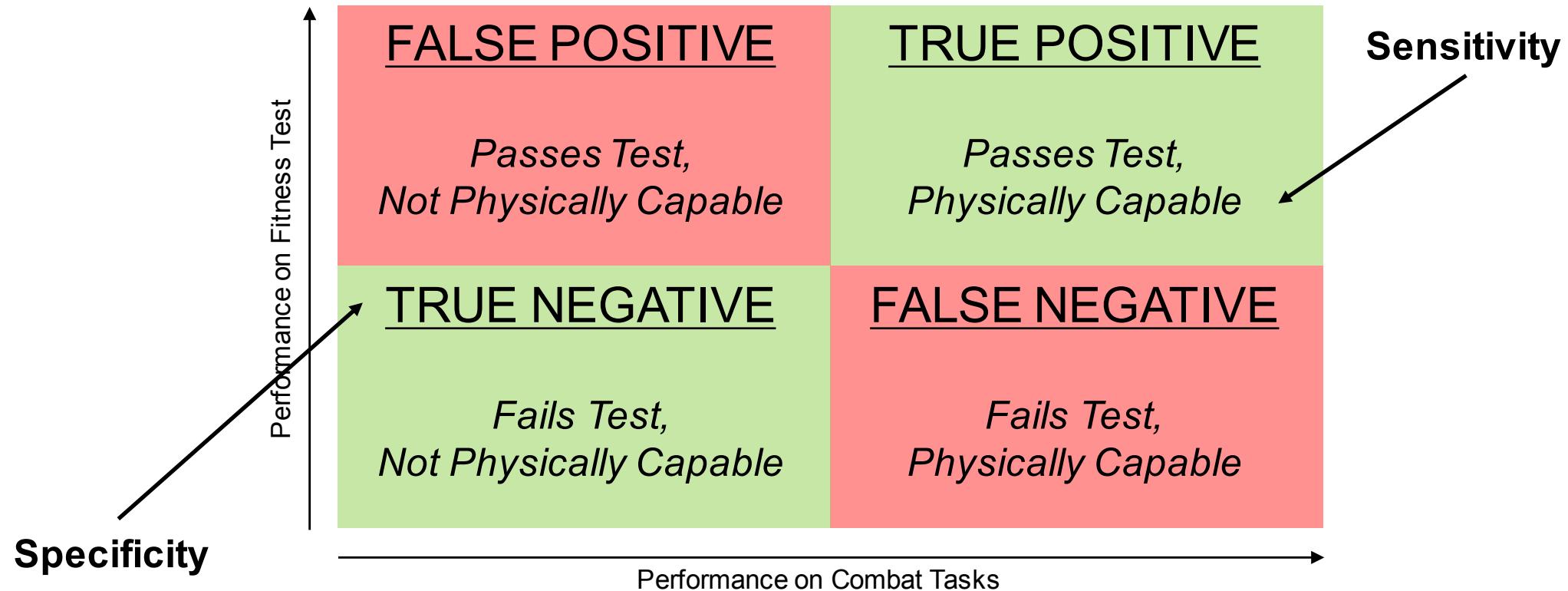
Occupationally Specific

# Classification Problem



# Two-Class Classification System

Goal is to build a fitness test that passes all individual's physically capable and fail all individuals not physically capable of combat tasks.



## The Youden Index

$$J = \max\{sensitivity(\theta) + specificity(\theta) - 1\}$$

# Validation

Compare new scale method  
to common STEN method to  
ensure performance is not  
degraded

Confirm classification  
performance

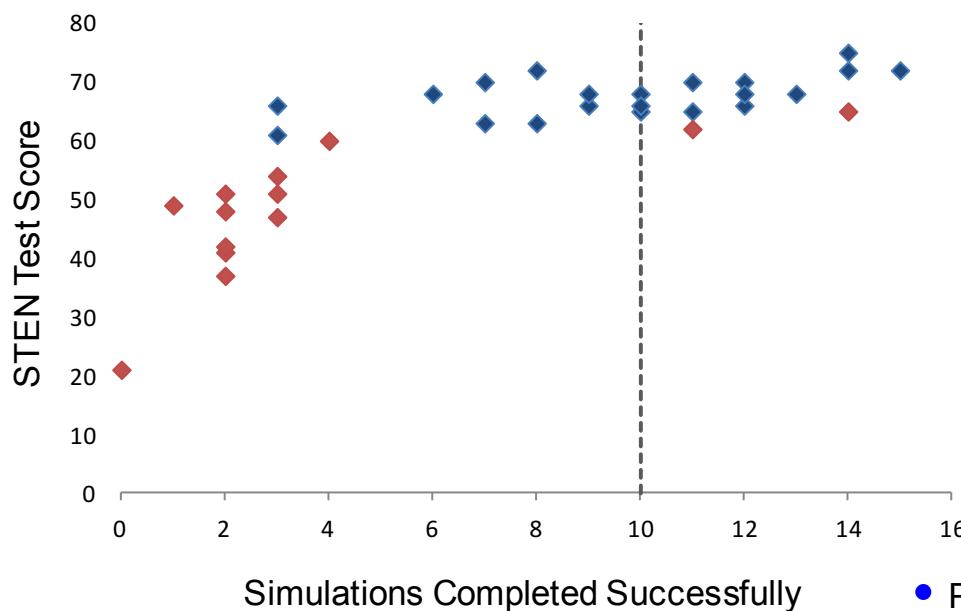


# Assessing Fitness Tests' Classification Accuracy

## STEN Scale Based Test Results

$\hat{J}$  Main Data = .79

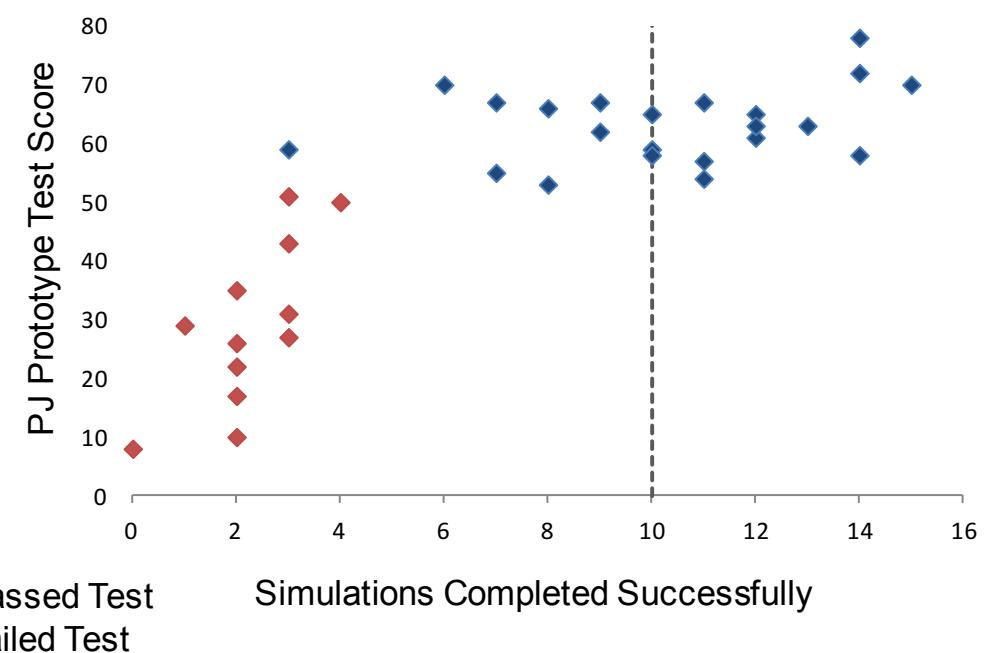
$\hat{J}$  Validation Data = 0.41  
95% Conf. Interval: [0.20, 0.70]



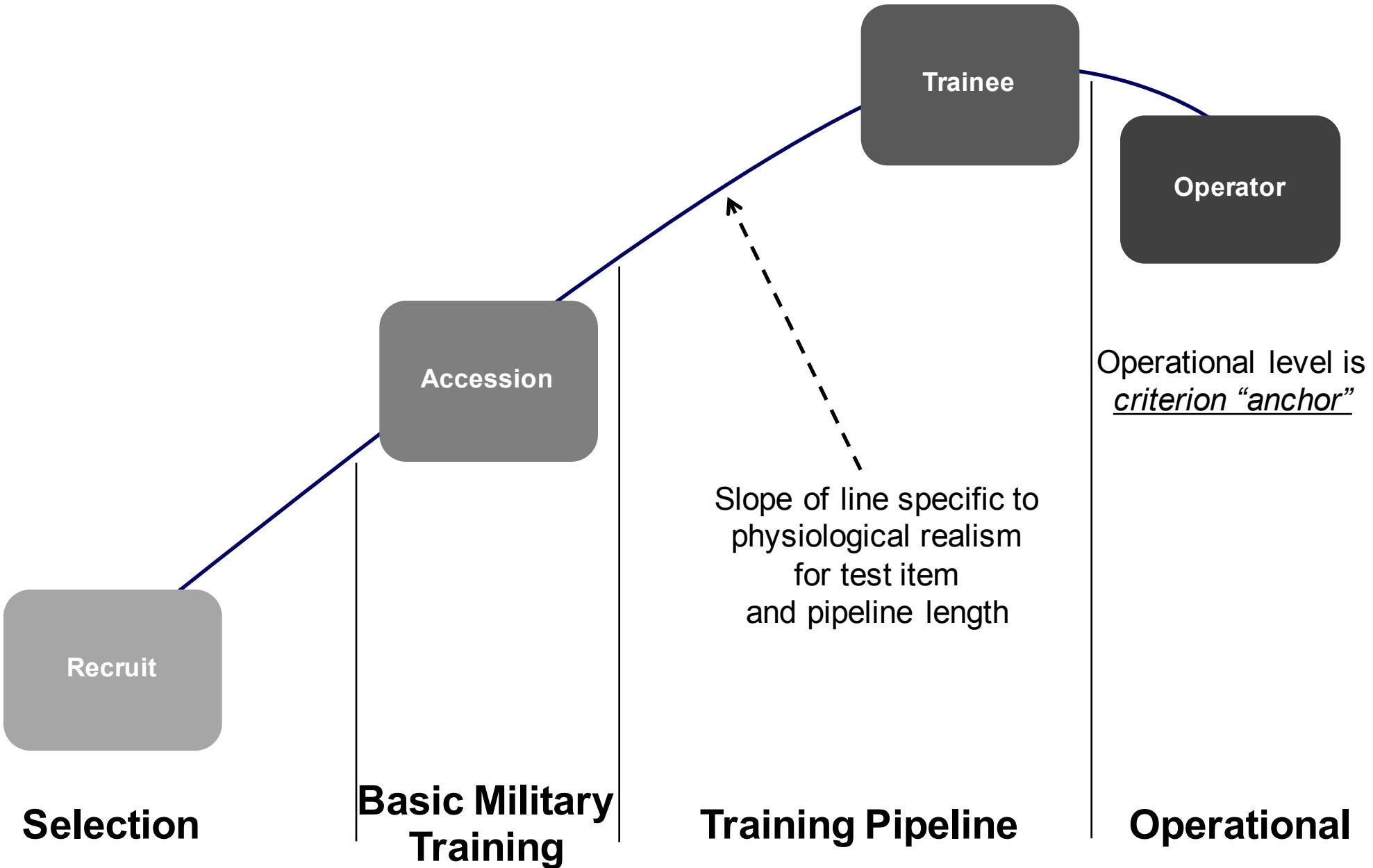
## Prototype Test Results

$\hat{J}$  Main Data = 0.78

$\hat{J}$  Validation Data = 0.60  
95% Conf. Interval: [0.35, 0.89]



# Standards for all BA Levels



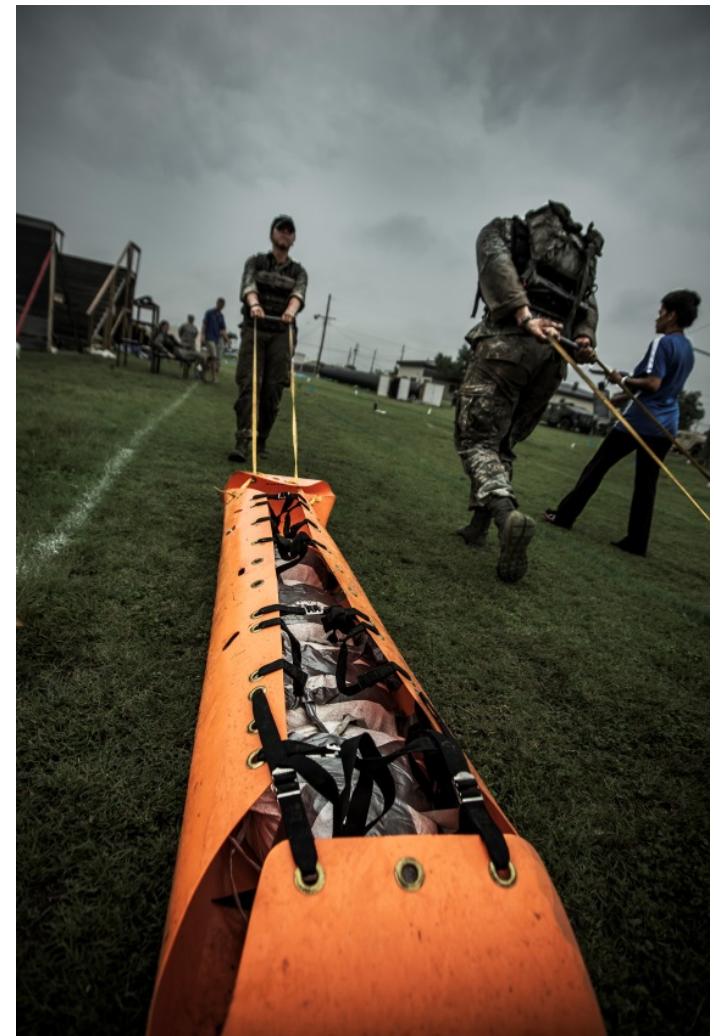
# Summary

Developed prototype BA fitness tests

- Unique to each career field
- Predictive of occupational performance
- Gender-neutral
- 4 levels of tests for full accession pipeline

New method for occupational test scales

- Better link between scale & task performance
- Classification performance consistent with accepted STEN approach



# Questions?



WISR & Public Law:

Batterton, K., Hale, K., Murphy, E., Women on the Battlefield: Data, Science, and the Law., *The Strategy Bridge*, 21 April 2016, <http://thestrategybridge.com/the-bridge/2016/4/21/women-on-the-battlefield-data-science-and-the-law>.

# AF WISR Fitness Standards Team

Air Force Fitness Standards and Testing Unit

*Dr. Neal Baumgartner*

HAF A3O

ACC & AFSOC

AETC SAS

RAND

