SECTION F

POTENTIAL TEST BATTERY VALIDATION

This study emphasized <u>construct and criterion validity</u>. The basic assumption was that underlying physiological readiness variables are the physical performance capabilities that should be addressed in a test battery. The various analyses were used to identify a test battery (i.e., a particular cluster of tests) that has construct and criterion validity. The job analysis identified essential functions and physical tasks of the job. Subject matter experts (SME) from the 19 agencies verified those functions and developed three specific task scenarios comprised of the most critical of those functions. The combination of the job-task analysis and the SME verification established the content validity of the job-task scenarios. These job-task scenarios became the %criterion-referenced measurements+ to compare the predictor tests (the fitness tests) against.

The Thomas and Means consultant team analyzed the job-task scenarios to identify the components of fitness necessary to accomplish those tasks and to identify valid measurements of those fitness components, thus establishing construct validity of the fitness tests used in the study. To establish criterion validity for the fitness tests, the relationships between these tests and the content valid job-task tests were determined. We applied the following rationale for interpreting the data and selecting the test battery items, using the data as objective indicators of validity.

We employed a rationale of **economy of test administration.** That is, we analyzed the data to determine the fewest number of test items that accounted for the most variance of performance. We accomplished this using a narrowing process. First, the job analysis data suggested an initial test battery that would validly measure underlying fitness dimensions. Secondly, a narrowing process relying on physical

performance data identified the <u>least</u> number of test items that accurately characterized the ability to do the job.

The narrowing and selection of test items followed a chain of logic based upon the various statistical analyses. There were three steps to define the battery in terms of construct and criterion validity: 1) conclusions from job analysis and job requirement data to insure the fitness tests measured the underlying fitness factors necessary to perform essential physical job tasks, 2) interpretations resulting from analyzing the relationships between the fitness and job-task test scores, and 3) interpretations from the specificity and sensitivity analysis.

The first step aids in insuring that the potential fitness test battery items have some construct validity for being underlying dimensions that are job related. The second step insures that the potential fitness test items have statistically predictive relationships to the job task simulation tests. However, those relationship data only document a relationship between the fitness test scores and the job task simulation test scores at a general level. It does not provide the data that denotes what specific score (or cutpoint) predicts the criterion cutpoint on the job task simulation test. The specificity and sensitivity analysis provide those data.

This Section details the process for defining a **potential fitness battery** that has criterion validity as a predictor of performance on the job task simulation test battery at a generic level. **Potential battery** is the term applied because the final definition of the test battery is obtained through the specificity and sensitivity analysis described in the next Section. That analysis looks at the predictability of specific fitness tests score cutpoints for predicting specific criterion performance on the job task simulation tests.

We looked at different sources of data to define and verify a common core of fitness factors or constructs. We applied the following assumptions to that judgment

process:
- That the predictive tests are those tests that measure those fitness areas that the job-task analysis and supporting data indicated were important for the job.

- The correlational and regression data provide direction for defining a battery of fitness tests that are related to the various physical aspects of the job.

Throughout all these steps we attempted to reduce the potential test battery to those tests that are independent and that do not duplicate measurement.

Results

1. JOB ANALYSIS DATA

The physical fitness tests do <u>not</u> demonstrate content validity in that the test items are not job-task simulations. However, significant correlations exist between those fitness test items and the job-task simulations, which do have content validity. The specific job-task test items were rated frequent and critical. In turn, the Thomas and Means consultant team defined the underlying variables of those tasks. The use of force critical incident and injury/absenteeism review also implied certain underlying fitness areas were important. The job data clearly suggests the following fitness factors (the tests measuring those factors are in parentheses) as the underlying physical factors for officers ability to perform the physical tasks of the job:

Aerobic power (1.5-mile run)
Anaerobic power (300-meter run)
Upper body muscular endurance (push-up)
Upper body strength (1RM bench press - raw score or ratio score)
Trunk endurance (sit-up)
Flexibility (sit and reach)
Leg power (vertical jump)
Agility (Illinois agility run)
Body composition (% fat)

In summary, the job analysis and supporting job data indicate that the physical fitness areas are essential for performing the job and can be classified as underlying variables of the content valid job tasks.

2. RELATIONSHIP ANALYSIS - UNIVARIATE CORRELATIONS BETWEEN VARIABLES

One aspect of construct validity (if possible to establish) is criterion-related validity among variables. We established this by observing the intercorrelations between the physical fitness test items and the three job task scenario test scores. Those tests of physical fitness factors that demonstrated significant correlations with any of the job-task scenario items and their total score are as follows:

        1.5-Mile Run
        300-Meter Run
        Push-Up
        Sit up
        Illinois agility run
        1 RM bench press (raw score or ratio score)
        Vertical Jump

The statistically significant correlations indicate a measure of concurrent or predictive validity. That is, the fitness tests are predictive of performance on the job-task items.

3. REGRESSION ANALYSES

The regression analysis provides the strongest data for economically defining predictive fitness tests. As was previously mentioned, body fat was excluded from consideration. Those tests that appeared as predictive tests included the following:

1.5 mile run
Illinois agility run
300-meter run
1 RM bench press (raw score or ratio score)
1 minute sit up test
Push up

4. TEST BATTERY SELECTION

The judgment process to select the test battery based upon the statistical data sources followed a systematic decision making process. The results of the judgment process are presented in Table F1. Selecting the battery of tests required four steps:

1. Listing of all physical fitness test variables (n = 9).

2. Listing of criteria for test selection:

   a) The physical fitness test variables had to logically appear to be an underlying factor based on the job task analysis.

   b) The test had significant correlations with at least two of the job task scenario scores and the total score.

   c) The test was a significant predictor in at least two of the regression patterns.

3. We evaluated each fitness test item on each criteria.

4. To be considered for inclusion as part of the fitness battery a test had to have met at least two of the three criteria.

TABLE F1
TEST BATTERY SELECTION PROCESS
_____

CRITERIA

Test	Criteria a JTA	Criteria c Correlation	Criteria c Regression	Total

Test					Count
1.5-Mile Run	X		X	X	3
300-Meter	X		X		2
Push-Up	X		X		2
Sit-Up	X		X		2
Sit & Reach	X				1
Vertical Jump	X		X		2
1 RM bench press raw	X		X	X	3
1 RM bench press ratio	X	X			2
Illinois agility run	X		X	X	3

Using these criteria, eight items were eligible for battery inclusion (1.5-mile run, 300-meter run, push-up, sit-up, vertical jump, 1 RM bench press raw score or ratio score and Illinois agility run). Using this judgment process these eight specific test items emerge as tests demonstrating construct and criterion validity based upon the statistical data.

Based upon this judgment process, a **<u>seven-item</u>** fitness test battery consisting of the following fitness tests could serve as the potential fitness battery:
1. 1.5-Mile Run
2. 300-Meter Run
3. Push-ups
4. Sit-ups
5. Vertical Jump
6. 1 RM bench press (raw score or ratio score)
7. Illinois agility run

The specificity and sensitivity analyses, reported in the next Section, provides the data from which to finalize the items in the fitness test battery and the specific score cutpoints that are predictive of criterion performance.