ANALYSIS OF INTER-RELATIONSHIPS BETWEEN SOME ARMED FORCES PHYSICAL FITNESS TESTS

CURTIS ROY RYLANDER *

IT IS THE PURPOSE of this study to determine, by statistical analysis, the inter-relationships between the several armed forces physical fitness tests and test items in order to establish the validity of various combinations of these test items in predicting physical fitness.

During World War II the armed forces adopted the term 'physical fitness'; the concept of physical fitness agreed upon, for all practical purposes, was the capacity to work hard over a long period of time without appreciable loss of efficiency. This concept was interpreted by the various agencies as motor fitness, which included the ingredients of muscular strength, power and endurance, cardio-respiratory endurance, agility, coordination, flexibility, and speed.[1]

The similarity of the meaning of the terms 'motor fitness' and 'physical fitness' can be readily seen in the definition of the former as stated by Brock, Cox, and Pennock: [2]

> Motor fitness is the final criterion through which all other elements of physical fitness are seen and measured in man. How continuously and efficiently he performed his daily work in industry, on the farm, in the armed forces or in athletic performances, was at one time the only criterion that man had of physical fitness. He might know little or nothing about scientific facts of body structure, or functioning of the organs, strength test on the dynamometer, or organic efficiency tests but he could understand an outstanding performance displaying power, speed and endurance.

---

* Department of Physical Education. This article is a summary of a master's thesis submitted in the School of Education of New York University, 1949.

[1] L. A. Larson, "Some Implications of the War Physical Fitness Program for Schools and Colleges," *First Year Book of the Eastern District Association*, American Association for Health, Physical Education and Recreation (1947), p. 54.

[2] J. D. Brock, W. A. Cox, and E. W. Pennock, "Physical Fitness," *Research Quarterly, Supplement* (May 1941), Ch. VII, pp. 407-415, as quoted by T. K. Cureton, *et al., Physical Fitness Appraisal and Guidance*, St. Louis, 1947, p. 390.

The development of tests for rating physical condition and efficiency has been in progress for a number of years; however, it was during World War II that this development received particular emphasis.[3] Each of the branches of the armed service constructed and utilized physical tests as an essential phase of its physical fitness program.

The armed forces physical fitness tests were particularly adaptable to military operations because they could be set up readily within the existing framework on an objective basis. This characteristic is equally applicable to the schools and colleges. However, because of the immediate need for the development of a test which would evaluate status and progress for the armed forces personnel, these tests (with the exception of the Army Air Forces Physical Fitness Test) were constructed in a minimum amount of time and without the opportunity for prolonged investigation. Thus the validity and reliability of these tests and test items may still be questioned, and if any further use of these tests is to be made, a more careful analysis should be conducted under more acceptable conditions, with sufficient time to complete the investigation satisfactorily.

An investigation is proposed, therefore, in the following two areas:

1. A determination of the validity of the various armed forces physical fitness tests and test items.

2. A development of possible new combinations of test items for predicting physical fitness through a study of the inter-relationships between the various test items contained in the armed forces tests and the criterion.

This study is confined to an analysis of the following official armed forces physical fitness tests: (1) the Army Air Forces (Outdoor) Physical Fitness Test, (2) the Army Ground Force (Indoor and Outdoor) Physical Fitness Test, (3) the Navy Standard Physical Fitness Test, (4) the Army Specialized Training Program Physical Efficiency Test, and (5) the Tests of Physical Fitness in Naval Aviation.

The criterion of performance is similar to the measure of physical

[3] H. H. Clarke, *The Application of Measurement to Health and Physical Education*, New York, 1946, p. 174.

fitness as established by Larson in the construction of the Springfield College Test,[4] and later in the construction of the Army Air Forces Physical Fitness Test.[5] The criterion consists of fifteen motor ability tests, and two static dynamometrical strength tests. It is designed to measure approximately the following aspects of physical fitness: endurance, speed, arm and shoulder girdle strength, abdominal strength, forearm strength, explosive power, agility, balance, and sensory-motor coordination. The constituents and tests of the criterion of performance are as follows:

1. Endurance—Endurance Index (time for 60-Yard Dash times 6) minus (time for 360-Yard Run)
2. Speed—60-Yard Dash
3. Speed and Endurance
   a. 360-Yard Run
   b. 360-Yard Shuttle Run (60-yard shuttle distance)
4. Muscular Endurance (Arm and Shoulder Girdle Strength)
   a. Chinning
   b. Floor Push-Ups
   c. Dipping
5. Muscular Endurance (Abdomen)
   a. Sit-Ups
   b. Leg Lifts
6. Muscular Strength (Forearm)
   a. Right Grip
   b. Left Grip
7. Explosive Power
   a. Vertical Jump
   b. Rope Climb
   c. Three Standing Broad Jumps
8. Agility—Burpee (20-second)
9. Body Coordination
   a. Cozens Dodge Run
   b. Baseball Throw for Distance

Research in the construction of short tests for predicting physical

[4] L. A. Larson, "A Factor and Validity Analysis of Strength Variables and Tests with a Test Combination of Chinning, Dipping, and Vertical Jump," *Research Quarterly* (December 1940), p. 82.

[5] L. A. Larson, "Some Findings Resulting from the Army Air Forces Physical Training Program," *Research Quarterly* (May 1946), p. 144.

fitness has received a great deal of attention since the beginning of World War II in 1939.

Larson,[6] in the construction of the Springfield College Test, found that the three variable—chinning, dips, and vertical jump—provided the best combination of strength variables for the prediction of a 15-variable criterion of motor ability.

In a later study conducted by Larson,[7] a more comprehensive statistical analysis was made of thirty-three motor ability variables and tests. One of the major purposes of this research was to utilize the information gained from the statistical analysis to construct a simple and valid outdoor test and indoor test which could effectively predict total motor ability, as described by the 25-variable motor ability criterion used.

Bookwalter utilized the index—chins plus push-ups times vertical jump—in an effort to establish simple, valid, and economical tests of motor fitness.[8] This simplification of previously established tests was validated against a 20-item criterion of performance and later against a more select 12-item criterion. In the interests of even greater adaptability, straddle chinning and standing broad jump were substituted in the original index (for chins and vertical jump, respectively), making possible a total of four indices to be used in predicting physical fitness.

Cureton, Welser, and Huffman,[9] in an attempt to develop a short motor fitness test, conducted an analysis of the University of Illinois (14-Item) Motor Fitness Test. This 14-item test was derived from correlations with a 30-item criterion, which was formulated on the basis of five items of varying difficulty in each of the six categories of balance, flexibility, strength, agility, power, and endurance. As a result of this analysis, a 7-Item Short Test was recommended.

Philips[10] developed the (3-item) JCR Test at Randolph Field, Texas, in 1941 in response to the request for a fitfiness test which would be practical and suitable to all levels of the Physical Training

[6] Larson, *Research Quarterly* (December 1940), *loc. cit.*

[7] Larson, "A Factor Analysis of Motor Ability Variables and Tests, with Tests for College Men," *Research Quarterly* (October 1941), p. 499.

[8] K. W. Bookwalter, "Test Manual for Indiana Motor Fitness Indices for High School and College Age Men," *Research Quarterly* (December 1943), p. 356.

[9] T. K. Cureton, Jr., L. Welser, and W. J. Huffman, "A Short Test for Predicting Motor Fitness," *Research Quarterly* (May 1945), p. 106.

[10] B. E. Phillips, "The JCR Test," *Research Quarterly* (March 1947), p. 12.

Program for Air Corps Cadets. Larson's Springfield College Test [11] was chosen as the test most nearly satisfying the requirements placed upon it. This test was modified to produce the JCR Test, which on *a priori* grounds, would appear to measure, in the most practical manner, the type of fitness which the air corps program was attempting to develop in its air crew personnel.

The Army Air Forces Physical Fitness Test,[12] which consists of chins, sit-ups, and 300-yard shuttle run (60-yard shuttle distance), was developed from a 15-variable criterion similar to the one used in the construction of the Springfield College Test.[13] Furthermore, in order to make the administration of this test possible indoors, an experiment was conducted on the selection of an indoor shuttle run. As a result of this analysis, the 250-yard run of 25-yard shuttle distance was selected for the indoor battery.

The Armed Forces Physical Fitness Tests

A survey of the literature and the War Deparment and Navy Department training manuals revealed seven official armed forces physical fitness tests:

1. The Navy Standard Physical Fitness Test.[14]
2. Tests of Physical Fitness in Naval Aviation.[15]
3. The Army Ground Force Physical Fitness Test (indoor battery and outdoor battery).[16]
4. The Army Air Forces Physical Fitness Test (indoor battery and outdoor battery).[17,18]
5. The Army Specialized Training Division Physical Efficiency Test.[19]

[11] Larson, *Research Quarterly* (December 1940), *loc. cit.*

[12] Larson, *Research Quarterly* (May 1946), *loc. cit.*

[13] Larson, *Research Quarterly* (December 1940), *loc. cit.*

[14] Training, Physical Fitness Test, *Physical Fitness Manual for the U. S. Navy.* Washington, D. C.: Office of the Chief of Naval Operations, U. S. N., Training Division, Physical Section, 1943, Chap. IV.

[15] Training, Physical Fitness Test, *Tests of Physical Fitness in Naval Aviation.* Washington, D. C.: Office of the Chief of Naval Operations, U. S. N., Aviation Training Division, January 1946. 28 pp.

[16] Training, Physical Fitness Test, *Physical Training.* Washington, D. C.: Headquarters Army Ground Force, War Department—Field Manual 21-20, January 1946, Chap. 17.

[17] Larson, *Research Quarterly* (May 1946), *loc. cit.*

[18] Training, Physical Fitness Test, *War Department—Army Air Forces Regulation No. 50-10.* Washington, D. C.: Headquarters Army Air Forces, March 21, 1945.

[19] Training, Physical Fitness Test, *Program of Physical Training.* Washington,

Except in the case of the Army Air Forces Physical Fitness Test, no statistical evidence has been reported for the reliability, objectivity, or validity of these test batteries.

Procedure

The data employed in this study were obtained from a representative sampling of the male student body at the University of Delaware in Newark, Delaware. The total number of students in the sample was 144, of whom 38 were athletes (members of freshmen or varisity sports squads) and 106 were non-athletes. The age range was from 17 to 29 years.

The various physical fitness tests and the 17-item criterion of performance were administered during the fall of 1947 in the regularly scheduled classes. The administration of the tests was completed in a period of approximately eleven weeks, and was preceded by two weeks of intensive physical conditioning, in order to prevent the students from being handicapped by lame muscles during the actual testing.

In administering the various armed forces test batteries, use was made, wherever possible, of those alternate test items which required a minimum of equipment and afforded greater ease in administration. This was done in the following test batteries:

1. Tests of Physical Fitness in Naval Aviation: the agility run replaced the speed-agility test.

2. The Army Ground Force (Indoor) Physical Fitness Test: the 60-second squat thrust replaced the indoor shuttle run.

In the case of the Army Air Forces Physical Fitness Test, only the outdoor battery was used in this study. The indoor battery was not included because the lack of space and equipment required for the indoor shuttle run precluded its use.

Before the administration of each test battery, the investigator gave detailed oral instructions in, and demonstrated the correct performance of, each test item to the students being tested. In every case the instructions given were taken from the original source of each test.

A total of six professional students of physical education were carefully trained to assist in the conduct of the tests as examiners. This was done with the assistance of a colleague of the investigator

D. C.: Headquarters Army Ground Force, Service of Supply, Army Specialized Training Program, March 12, 1943.

in the Department of Physical Education at the University of Delaware, Professor D. Kenneth Steers, who also assisted in the conduct of the tests as an examiner. The investigator supervised the testing procedures and was present at all times during the conduct of the tests.

In the statistical analysis, standard scores were utilized in the computation of all correlation coefficients, except in the case of the reliability coefficients, where the raw scores were used. The use of standard scores, in the place of raw scores, is justifiable from the standpoint that " standard or Z-scores . . . have the *same* form of distribution as the original scores, and are simply original scores expressed in sigma-units." [20]

Reliability coefficients were computed for each of the twenty-six different items which were represented in one or more of the armed forces physical fitness test batteries and the 17-item criterion. In every case where a test item was repeated for the purpose of determining the reliability of that item, the retest was given within two weeks from the time of the initial test.

The following is a brief summary of the procedures and techniques used in the statistical analysis of each of the armed forces physical fitness tests:

1. A matrix of zero-order correlations was computed to obtain the relationships of each variable with each of the other variables in the battery and the composite score of the battery.

2. A zero-order correlation of each test item and the test battery's composite score with the composite criterion score was computed to obtain the relationships of the test battery and its individual test items to the criterion of performance.

3. The multiple correlation coefficient of the test battery with the criterion, and the beta weights for each of the items in the test battery, were computed utilizing the Doolittle multiple correlation technique,[21] and the Wherry-Doolittle test selection method.[22,23]

In order to determine possible new test combinations for the

[20] H. E. Garrett, *Statistics in Psychology and Education*, New York, 1947, p. 157.
[21] C. C. Peters, and W. R. Van Voorhis, *Statistical Procedures and Their Mathematical Bases*, New York, 1940, p. 220.
[22] Garrett, *op. cit.*, p. 435.
[23] W. H. Stead, C. L. Shartle, *et al.*, *Occupational Counseling Techniques*, New York, 1940, Appendix V.

prediction of physical fitness, the twenty-six different test items represented in one or more of the test batteries and the criterion were grouped into the various constituents of physical fitness. This grouping of test items is based upon previously made studies.[24,25,26]

Zero-order intercorrelations were then computed for the tests within each constituent. After a careful analysis of these intercorrelations and the criterion correlations for each item, several possible test combinations were selected and further analyzed utilizing the Wherry-Doolittle test selection method. In addition, the multiple regression (standard score) equations and the standard error of estimate ($\sigma_{(est.\ \bar{z}_0)}$) for each regression equation were computed for these new test combinations, and also for the naval aviation test battery and a modification of the Army Specialized Training Program Physical Efficiency Test.

All statistical work involved in this study was done by the investigator; all calculations were repeated.

Interpretation of Data

The rank order in which the various armed forces physical fitness tests predict the criterion of performance is as follows:

1. Tests of Physical Fitness in Naval Aviation ($\bar{R}_0 = .7440$)
2. The Army Specialized Training Program Physical Efficiency Test ($\bar{R}_0 = .7307$)
3. The Army Ground Force (Outdoor) Physical Fitness Test ($\bar{R}_0 = .6660$)
4. The Army Ground Force (Indoor) Physical Fitness Test ($\bar{R}_0 = .6527$)
5. The Navy Standard Physical Fitness Test ($\bar{R}_0 = .5941$)
6. The Army Air Forces (Outdoor) Physical Fitness Test ($\bar{R}_0 = .5929$)

All of the corrected multiple correlation coefficients of the test batteries, as indicated above, are significant at the one per cent level of confidence with (N — m) degrees of freedom, wherein 'N'

[24] Cureton, *et al.*, *Physical Fitness Appraisal and Guidance*, p. 392.

[25] Cureton, "Endurance of Young Men" (Analysis of Endurance Exercises and Methods of Evaluating Motor Fitness), *Monographs of the Society for Research in Child Development*, Vol. X, Ser. No. 40, No. 1, Society for Research in Child Development, National Research Council, 1945, p. 3.

[26] Larson, *Research Quarterly* (October 1941), p. 502.

equals 144, the number of students tested, and 'm' equals the number of variables in the test battery.[27]

The 5-item naval aviation test battery, whose corrected multiple correlation coefficient equals .7440, is suitable only for group prediction purposes (see Tables I, II, III).[28] In the interests of administrative economy, the step test item in this battery may well be eliminated without decreasing substantially the validity of the battery. The corrected multiple correlation coefficient for the remaining four test items is .7429. The test items and the standard score equations for the four and five item combinations are as follows:

1. Chinning
2. Sit-Ups (with wand)
3. Vertical Jump
4. Agility Run
5. Step Test

$\bar{R}_0\,1, 2, 3, 4 = .7429$; $\sigma_{(\text{est. } \bar{Z}_0)} = 326.67$

$\bar{Z}_0 = .1462\,Z_1 + .2084\,Z_2 + .4339\,Z_3 + .2818\,Z_4$

$\bar{R}_0\,1, 2, 3, 4, 5 = .7440$; $\sigma_{(\text{est. } \bar{Z}_0)} = 326.07$

$\bar{Z}_0 = .1360\,Z_1 + .2081\,Z_2 + .4286\,Z_3 + .2769\,Z_4 + .0714\,Z_5$.

The 7-item Army Specialized Training Program Physical Efficiency Test battery has greater validity with the elimination of two of its test items, namely, pull-ups and squat jumps. The corrected multiple correlation coefficient for the remaining 5-item test battery is .7342, as compared with .7307, the corrected multiple correlation coefficient for the 7-item test battery (see Table VI). The use of either of these two test batteries is limited to group surveys only. The use of the shorter and more valid 5-item test battery, however, is recommended over the use of the original 7-item test battery. The 5-item test battery and its standard score equation are as follows:

[27] Garrett, *op. cit.*, p. 426.

[28] All test batteries in this study are judged upon the following acceptable levels of validity, reliability, and objectivity:

.85 and up . . Adequate for individual measurement.

.75 — .84 . . Adequate for group measurement, but not satisfactory for individual measurement.

below .75 . . Entirely inadequate for individual measurement, but useful for group averages and surveys.

1. 300-Yard Shuttle Run (150-yard shuttle distance)
2. Floor Push-Ups
3. 100-Yard Pick-a-back
4. Sit-Ups
5. Burpee (20-second)

$\bar{R}_0\ 1, 2, 3, 4, 5 = .7342$; $\sigma_{(est.\ \bar{Z}_0)} = 331.32$

$\bar{Z}_0 = .3217\ Z_1 + .2836\ Z_2 + .2314\ Z_3 + .1073\ Z_4 + .0863\ Z_5$

The 5-item Army Ground Force (Outdoor) Physical Fitness Test battery has greater validity with the elimination of squat jumps from the battery. The corrected multiple correlation coefficient for the 5-item test battery is .6660, while the corrected multiple correlation coefficient for the 4-item test battery (5-item test battery minus squat jumps) is .6675 (see Table XII). These test batteries are suitable for group survey purposes only, and are not recommended by the investigator. (Beta weightings for the 5-item test battery are reported in Table XI.)

The 5-item Army Ground Force (Indoor) Physical Fitness Test battery has a corrected multiple correlation coefficient of .6527 (see Table IX), and is satisfactory for group survey purposes only. The investigator does not recommend this test battery. (Beta weightings for this test battery are reported in Table VIII.)

The corrected multiple correlation coefficient for the 5-item Navy Standard Physical Fitness Test is .5941. However, with the elimination of the squat jumps test item from this battery, the corrected multiple correlation coefficient increases to .5979 for the remaining four test items (see Table XV). These test batteries are adequate for group survey purposes only, and are not recommended by the investigator. (Beta weightings for the 5-item test battery are reported in Table XIV.)

The Army Air Forces (Outdoor) Physical Fitness Test battery has a corrected multiple correlation coefficient of .5929 (see Table XVIII), and is suitable for group survey purposes only. The investigator does not recommend this test battery. (Beta weightings for this test battery are reported in Table XVII.)

The twenty-six different variables represented in one or more of the armed forces test batteries and in the criterion of physical fitness performance have varying degrees of validity and reliability (see Table XIX).

As indicated by the data presented in this study, no single test

item has a sufficiently high relationship to the physical fitness criterion so that it alone could be said to be an adequate measure of the fitness of an individual.

The following test items have little or no relationship to the performance criterion, have a low net relative contribution to the prediction of the criterion in their respective test batteries, and show little relationship to the other test items in their respective test batteries:

1. Endurance Index
2. Cozens Dodge Run
3. Step Test
4. Squat Jumps
5. Burpee (20-second)
6. Burpee (60-second).

However, all the test items which enter into this study show some relationship to the criterion, and are significant at the one per cent level of confidence.[29] It may be surmised, therefore, that each measures a part of total physical fitness.

The reliability coefficients of the following test items are questionably low:

1. Endurance Index (.6158)
2. 60-Yard Dash (.7144)
3. 300-Yard Shuttle Run (60-yard shuttle distance) (.5643)
4. Floor Push-Ups (.7338)
5. Sit-Ups (.6045)
6. Burpee (20-second) (.4690)
7. Burpee (60-second) (.5553).

The exact reason for these low reliability coefficients is not clearly understood by the investigator. Logical causes, however, would seem to come under the heading of 'chance errors.' These chance errors are as follows:

1. Effects of practice and conditioning.
2. Fluctuations in interest and attitude.
3. Fluctuations in health.
4. External environmental changes.

These influencing factors undoubtedly affected the reliability of

[29] Garrett, *op. cit.*, p. 299.

the various test items, especially since the testing period was quite long (eleven weeks).

The reliability of certain test items is also affected to some degree by the amount of, or lack of, control within the test items themselves. The reliability of sit-ups, for example, increases substantially when a time limit is imposed, or when, as in the case of sit-ups in the naval aviation test battery, a wand is used by the testee as a part of the test. As indicated by the results obtained in this study, the reliability of sit-ups is .6045, the reliability of sit-ups (2-minute) is .8341, and the reliability of sit-ups (with wand) is .8443. The use of the latter two variations of the sit-ups test is therefore recommended, not only because of their greater reliability, but also because of the greater ease and economy in administration which they afford.

New Combinations of Test Items

The following four 5-item combinations of test items for the measurement of physical fitness are recommended for further study. These test combinations are a result of the detailed study of the twenty-six variables represented in the various armed armed forces test batteries and the criterion (see Table XIX).

The corrected multiple correlation coefficients of these test batteries are significant at the one per cent level of confidence,[30] and range from .8238 to .9921.

Test I

This test battery is recommended as adequate for group measurement; it is not satisfactory for individual prediction purposes. The maximum multiple correlation coefficient obtained is .8238 (see Table XXI). The test battery and its standard score equation are as follows:

1. Floor Push-Ups
2. Left Grip
3. 60-Yard Dash
4. Sit-Ups (2-minute)
5. 100-Yard Pick-a-back

$R_0\ 1, 2, 3, 4, 5 = .8238$; $\sigma_{(est.\ Z_0)} = 276.64$

$Z_0 = .2479\ Z_1 + .4132\ Z_2 + .2143\ Z_3 + .2027\ Z_4 + .1808\ Z_5$

[30] Garrett, *op. cit.*, p. 426.

Test II

This test battery is recommended as adequate for individual measurement (see Table XXII). The test battery and its standard score equation are as follows:

1. Baseball Throw for Distance
2. Floor Push-Ups
3. Left Grip
4. 300-Yard Shuttle Run (150-yard shuttle distance)
5. Sit-ups (2-minute)

$\bar{R}_0\,1, 2, 3, 4, 5 = .9402$; $\sigma_{(\text{est. } \bar{Z}_0)} = 166.22$

$\bar{Z}_0 = .4938\,Z_1 + .2309\,Z_2 + .3326\,Z_3 + .2334\,Z_4 + .1027\,Z_5$

Test III

This test battery is recommended as adequate for individual measurement (see Table XXIII). The test battery and its standard score equation are as follows:

1. Baseball Throw for Distance
2. Three Standing Broad Jumps
3. Left Grip
4. Sit-Ups (2-minute)
5. Floor Push-Ups

$\bar{R}_0\,1, 2, 3, 4, 5 = .9701$; $\sigma_{(\text{est. } \bar{Z}_0)} = 118.44$

$\bar{Z}_0 = .4858\,Z_1 + .3569\,Z_2 + .3313\,Z_3 + .1488\,Z_4 + .1435\,Z_5$

Test IV

This test battery is adequate for individual measurement (see Table XXIV). Furthermore, it is recommended by the investigator that this test battery be used as an unweighted index. It is realized that some of the accuracy is lost by doing so, but the saving in time and effort more than offsets this loss. Based on the sample employed in this study, this test battery's mean composite scores have a zero-order correlation of .8988 with the criterion. By using as scores either the unweighted means of the composite scores, or the unweighted composite scores, this test battery measures approximately 81 per cent of the variation in physical fitness within a group, and falls within the range of individual predictability.

The test battery and its standard score equation are as follows:

1. Baseball Throw for Distance

2. Three Standing Broad Jumps
3. 100-Yard Pick-a-back
4. 300-Yard Shuttle Run (150-yard shuttle distance)
5. Floor Push-Ups

$\bar{R}_0\ 1, 2, 3, 4, 5 = .9921$; $\sigma_{(est.\ \bar{Z}_0)} = 61.22$

$\bar{Z}_0 = .8547\ Z_1 + .8065\ Z_2 - .7477\ Z_3 + .2305\ Z_4 + .1206\ Z_5$

Summary

1. The various official armed forces physical fitness test batteries are adequate for group survey measurement only.

2. New test combinations for the prediction of physical fitness are herein reported. The investigator feels, however, that there is a need for further study and evaluation of the validity and reliability of these test batteries.

3. There is a need for the establishment of norms for the new physical fitness test combinations reported in this study, in order for it to be possible to interpret test scores adequately.

4. The step test, Burpee (20-second and 60-second), endurance index, Cozens dodge run, and squat jumps seem to be measuring different factors from the other test items in this study.

5. Further investigation is needed in order to determine more exactly what factors or traits our physical fitness tests are measuring.

6. There is a need for the study of possible improvement of physical fitness test items, in order to enhance their reliability.

TABLE I

MATRIX OF INTERCORRELATIONS FOR TESTS OF PHYSICAL FITNESS IN NAVAL AVIATION

	1	2	3	4	5	Naval Composite
Criterion	.4903	.4395	.6062	.4789	.2170	.5519
Naval Aviation Composite Score	.5872	.8321	.3908	.2952	.5730	
1. Chinning		.4605	.3503	.3413	.1930	
2. Sit-ups (with wand)			.2501	.1965	.1012	
3. Vertical Jump				.2450	.1420	
4. Agility Run					.1352	
5. Step Test						

TABLE II

ANALYSIS OF THE TESTS OF PHYSICAL FITNESS IN NAVAL AVIATION UTILIZING THE DOOLITTLE MULTIPLE CORRELATION TECHNIQUE

Items (In Standard Scores)	Beta Weightings	Contribution in Per Cent.	Criterion Correlation
1. Chinning	.1360	5.65	.4903
2. Sit-Ups (with wand)	.2081	13.23	.4395
3. Vertical Jump	.4286	56.13	.6062
4. Agility Run	.2769	23.43	.4789
5. Step Test	.0714	1.56	.2170

$\bar{R}_o = .7440$ *

* $\bar{R}_o$ equal the multiple correlation coefficient corrected for chance errors.

TABLE III

ANALYSIS OF THE TESTS OF PHYSICAL FITNESS IN NAVAL AVIATION UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\bar{R}_o$
1. Vertical Jump	.6062
2. Agility Run	.6929
3. Sit-Ups (with wand)	.7351
4. Chinning	.7429
5. Step Test	.7440

TABLE IV

MATRIX OF INTERCORRELATIONS FOR THE ARMY SPECIALIZED TRAINING PROGRAM PHYSICAL EFFICIENCY TEST

	1	2	3	4	5	6	7	A. S. T. P.* Composite
Criterion	.5738	.3458	.4215	.4237	.5367	.2310	.6214	.6042
A. S. T. P. Composite	.7330	.6388	.8321	.5367	.4799	.2058	.5687	
1. Push-Ups		.3301	.5215.	.6043	.3389	.1458	.4455	
2. Squat Jumps			.3391	.2003	.3067	.1834	.3604	
3. Sit-Ups				.3303	.2682	.0627	.3074	
4. Pull-Ups					.2418	.2059	.3591	
5. 100-Yard Pick-a-back						.1231	.5277	
6. Burpee (20-second)							.2117	
7. 300-Yard Shuttle Run (150-yard shuttle distance)								

* A. S. T. P. refers to the Army Specialized Training Program (Physical Efficiency Test).

TABLE V

ANALYSIS OF THE ARMY SPECIALIZED TRAINING PROGRAM PHYSICAL EFFICIENCY TEST UTILIZING THE DOOLITTLE MULTIPLE CORRELATION TECHNIQUE

Items (In Standard Scores)	Beta Weightings	Contribution in Per Cent	Criterion Correlation
1. Push-Ups	.2568	27.84	.5738
2. Squat Jumps	.0188	0.17	.3458
3. Sit-Ups	.1030	4.48	.4215
4. Pull-Ups	.0462	0.89	.4237
5. 100-yard Pick-a-back	.2296	22.26	.5367
6. Burpee (20-second)	.0795	2.66	.2310
7. 300-yard Shuttle Run (150-yard shuttle distance)	.3142	41.70	.6214

$\bar{R}_o = .7307$

TABLE VI

ANALYSIS OF THE ARMY SPECIALIZED TRAINING PROGRAM PHYSICAL EFFICIENCY TEST UTILIZING THE WHERRY-DOOLITTLE TEST SELELCTION METHOD

Test Items in Order of Selection	$\bar{R}_o$
1. 300-Yard Shuttle Run (150-yard shuttle distance)	.6214
2. Push-Ups	.7019
3. 100-Yard Pick-a-back	.7284
4. Sit-Ups	.7316
5. Burpee (20-second)	.7342
6. Pull-Ups	.7328
7. Squat Jumps	.7307

TABLE VII

MATRIX OF INTERCORRELATIONS FOR THE ARMY GROUND FORCE (INDOOR) PHYSICAL FITNESS TEST

	1	2	3	4	5	A. G. F.-I.* Composite
Criterion	.4509	—.0016	.5665	.4124	.3375	.5951
A. G. F.-I. Composite	.5818	.5333	.7412	.7313	.4468	
1. Pull-Ups		.1078	.5376	.2667	.3935	
2. Squat Jumps			.1458	.3223	.1923	
3. Push-Ups				.4097	.2427	
4. Sit-Ups (2-minute)					.1370	
5. Burpee (60-second)						

* A. G. F.-I. refers to the Army Ground Force (Indoor) (Physical Fitness Test)

TABLE VIII

ANALYSIS OF THE ARMY GROUND FORCE (INDOOR) PHYSICAL FITNESS TEST UTILIZING THE DOOLITTLE MULTIPLE CORRELATION TECHNIQUE

Items (In Standard Scores)	Beta Weightings	Contribution in Per Cent	Criterion Correlation
1. Pull-Ups	.1228	5.05	.4509
2. Squat Jumps	—.1918	12.31	—.0016
3. Push-Ups	.3728	46.49	.5665
4. Sit-Ups (2-minute)	.2612	22.81	.4124
5. Burpee (60-second)	.1997	13.34	.3375
	$\bar{R}_o = .6527$		

TABLE IX

ANALYSIS OF THE ARMY GROUND FORCE (INDOOR) PHYSICAL FITNESS TEST UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\bar{R}_o$
1. Push-Ups	.5665
2. Burpee (60-second)	.5991
3. Sit-Ups (2-minute)	.6251
4. Squat Jumps	.6484
5. Pull-Ups	.6527

TABLE X

MATRIX OF INTERCORRELATIONS FOR THE ARMY GROUND FORCE (OUTDOOR) PHYSICAL FITNESS TEST

	1	2	3	4	5	A. G. F.-O.* Composite
Criterion	.4353	.2287	.5390	.4526	.4951	.6345
A. G. F.-O. Composite	.5688	.5962	.6468	.7348	.6501	
1. Pull-Ups		.1475	.5354	.2274	.3892	
2. Squat Jumps			.1883	.2379	.1957	
3. Push-Ups				.3003	.3533	
4. Sit-Ups (2-minute)					.3349	
5. 300-Yard Shuttle Run (60-yard shuttle distance)						

* A. G. F.-O. refers to the Army Ground Force (Outdoor) (Physical Fitness Test)

TABLE XI

ANALYSIS OF THE ARMY GROUND FORCE (OUTDOOR) PHYSICAL FITNESS TEST UTILIZING THE DOOLITTLE MULTIPLE CORRELATION TECHNIQUE

Items (In Standard Scores)	Beta Weightings	Contribution in Per Cent	Criterion Correlation
1. Pull-Ups	.1089	5.14	.4353
2. Squat Jumps	.0478	0.99	.2287
3. Push-Ups	.3104	41.62	.5390
4. Sit-Ups (2-minute)	.2381	24.50	.4526
5. 300-Yard Shuttle Run (60-yard shuttle distance)	.2534	27.74	.4951

$\tilde{R}_o = .6660$

TABLE XII

ANALYSIS OF THE ARMY GROUND FORCE (OUTDOOR) PHYSICAL FITNESS TEST UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\tilde{R}_o$
1. Push-Ups	.5390
2. 300-Yard Shuttle Run (60-yard shuttle distance)	.6263
3. Sit-Ups (2-minute)	.6642
4. Pull-Ups	.6675
5. Squat Jumps	.6660

TABLE XIII

MATRIX OF INTERCORRELATIONS FOR THE NAVY STANDARD PHYSICAL FITNESS TEST

	1	2	3	4	5	Navy Composite
Criterion	.3021	.4000	.5694	.2000	.4126	.5391
Navy Composite	.4654	.8444	.8208	.5658	.4807	
1. Burpee (60-second)		.1834	.3701	.2379	.2779	
2. Sit-Ups			.5188	.2496	.1412	
3. Push-Ups				.2527	.5266	
4. Squat Jumps					.1885	
5. Pull-Ups						

TABLE XIV

ANALYSIS OF THE NAVY STANDARD PHYSICAL FITNESS TEST UTILIZING THE DOOLITTLE MULTIPLE CORRELATION TECHNIQUE

Items (In Standard Scores)	Beta Weightings	Contribution in Per Cent	Criterion Correlation
1. Burpee (60-second)	.0883	4.06	.3021
2. Sit-Ups	.1739	15.73	.4000
3. Push-Ups	.3501	63.85	.5694
4. Squat Jumps	.0137	0.10	.2000
5. Pull-Ups	.1765	16.25	.4126
	$\bar{R}_0 = .5941$		

TABLE XV

ANALYSIS OF THE NAVY STANDARD PHYSICAL FITNESS TEST UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\bar{R}_0$
1. Push-Ups	.5694
2. Pull-Ups	.5807
3. Sit-Ups	.5958
4. Burpee (60-second)	.5979
5. Squat Jumps	.5941

TABLE XVI

MATRIX OF INTERCORRELATIONS FOR THE ARMY AIR FORCES (OUTDOOR) PHYSICAL FITNESS TEST

	1	2	3	A. A. F.-O.* Composite
Criterion	.4744	.4108	.4231	.5129
A. A. F.-O. Composite	.9680	.5152	.4345	
1. Sit-Ups		.4372	.3250	
2. Pull-Ups			.1196	
3. 300-Yard Shuttle Run (60-yard shuttle distance)				

* A. A. F.-O. refers to the Army Air Forces (Outdoor) (Physical Fitness Test)

TABLE XVII

ANALYSIS OF THE ARMY AIR FORCES (OUTDOOR) PHYSICAL FITNESS TEST UTILIZING THE DOOLITTLE MULTIPLE CORRELATION TECHNIQUE

Items (In Standard Scores)	Beta Weightings	Contribution in Per Cent	Criterion Correlation
1. Sit-Ups	.2608	29.57	.4744
2. Pull-Ups	.2600	29.39	.4108
3. 300-Yard Shuttle Run (60-yard shuttle distance)	.3073	41.04	.4231

$\tilde{R}_o = .5929$

TABLE XVIII

ANALYSIS OF THE ARMY AIR FORCES (OUTDOOR) PHYSICAL FITNESS TEST UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\tilde{R}_o$
1. Sit-Ups	.4744
2. 300-Yard Shuttle Run (60-yard shuttle distance)	.5486
3. Pull-Ups	.5929

TABLE XIX

CRITERION CORRELATIONS, RELIABILITY COEFFICIENTS, AND INTERCORRELATIONS * FOR TWENTY-SIX VARIABLES

Constituent of Physical Fitness	Test Item	Criterion Correlation	Reliability
A. Endurance	Endurance Index	.2935	.6158
B. Speed	60-Yard Dash	.5644	.7144
C. Speed and Endurance	1. 360-Yard Run	.6071	.7451
	2. 360-Yard Shuttle Run (60-yard shuttle)	.5051	.8613
	3. 300-Yard Shuttle Run (60-yard shuttle)	.4951	.5643
	4. 300-Yard Shuttle Run (150-yard shuttle)	.6214	.8378
D. Muscular Endurance (Arm)	1. Chinning	.4903	.8263
	2. Floor Push-Ups	.5738	.7338
	3. Dipping	.5252	.8341
E. Muscular Endurance (Abdomen)	1. Sit-Ups	.4744	.6045
	2. Sit-Ups (2-minute)	.4526	.8341
	3. Sit-Ups (with wand)	.4395	.8443
	4. Leg Lifts	.4006	.8649

* The intercorrelations include only those correlations between the test items within each constituent group.

Constituent of Physical Fitness	Test Item	Criterion Correlation	Reliability
F. Muscular Endurance (Thigh and Leg)	Squat Jumps	.3458	.7658
G. Muscular Strength (Forearm)	1. Right Grip	.5345	.8405
	2. Left Grip	.5703	.8573
H. Explosive Power	1. Vertical Jump	.6062	.9135
	2. Rope Climb	.4811	.8992
	3. 3-Standing Board Jumps	.6657	.9212
I. Agility	1. Burpee (20-second)	.2310	.4690
	2. Burpee (60-second)	.3375	.5553
J. Body Coordination	1. Baseball Throw	.7452	.9356
	2. Cozens Dodge Run	.2935	.8249
	3. Agility Run	.4789	.6865
K. Total Body Strength, Speed, and Endurance	100-Yard Pick-a-back	.5367	.8847
L. Cardio-Respiratory Efficiency	Step Test	.2170	.7701

Intercorrelations:

C. r12 = .6730 r23 = .3832 r34 = .5056
r13 = .5521 r24 = .5885
r14 = .7058

D. r12 = .6043 r23 = .5626
r13 = .6807

E. r12 = .6213 r23 = .5434 r34 = .4540
r13 = .6192 r24 = .3955
r14 = .4405

G. r12 = .7094

H. r12 = .2976 r23 = .5403
r13 = .6362

I. r12 = .4517

J. r12 = .1795 r23 = .4244
r13 = .3794

TABLE XX

CRITERION CORRELATIONS AND INTERCORRELATIONS FOR A 12-ITEM PHYSICAL FITNESS TEST

	1	2	3	4	5	6	7	8	9	10	11	12	12-Item Composite
Criterion	.2935	.5644	.6214	.5738	.4526	.5703	.6057	.3375	.7452	.3458	.5367	.2170	.9195
1. Endurance Index		—.1101	.4764	.1755	.1692	—.0233	.1606	.1023	.1598	.1998	.2240	.0454	
2. 60-Yard Dash			.5661	.4188	.3440	.2141	.5695	.2821	.3864	.3296	.4872	.1990	
3. 300-Yd. Shuttle Run (150-Yd. shuttle)				.4455	.4513	.1814	.5254	.2607	.3604	.3604	.5277	.2679	
4. Floor Push-Ups					.3905	.2315	.4872	.1703	.2468	.3301	.3389	.0355	
5. Sit-Ups (2-minute)						.0483	.2660	.1370	.2832	.2499	.3288	.1319	
6. Left Grip							.1866	.1725	.2442	.0276	.2438	—.0050	
7. Three Standing Broad Jumps								.3688	.2830	.3197	.7523	.2119	
8. Burpee (60-second)									.1289	.2385	.3265	.0877	
9. Baseball Throw for Distance										.2237	.6026	.2721	
10. Squat Jumps											.3067	.2115	
11. 100-Yard Pick-a-back												.1413	
12. Step Test													

TABLE XXI

ANALYSIS OF FIVE-ITEM TEST BATTERY NUMBER ONE UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\tilde{R}_0$
1. Floor Push-Ups	.5738
2. Left Grip	.7267
3. 60-Yard Dash	.7848
4. Sit-Ups (2-minute)	.8107
5. 100-Yard Pick-a-back	.8238

Standard Score Equation:

$$\tilde{Z}_0 = .2479Z_1 + .4132Z_2 + .2143Z_3 + .2027Z_4 + .1808Z_5$$

TABLE XXII

ANALYSIS OF FIVE-ITEM TEST BATTERY NUMBER TWO UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\tilde{R}_0$
1. Baseball Throw for Distance	.7452
2. Floor Push-Ups	.8457
3. Left Grip	.9085
4. 300-Yard Shuttle Run (150-yard shuttle distance)	.9365
5. Sit-Ups (2-minute)	.9402

Standard Score Equation:

$$\tilde{Z}_0 = .4938Z_1 + .2309Z_2 + .3326Z_3 + .2334Z_4 + .1027Z_5$$

TABLE XXIII

ANALYSIS OF FIVE-ITEM TEST BATTERY NUMBER THREE UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\tilde{R}_0$
1. Baseball Throw for Distance	.7452
2. Three Standing Broad Jumps	.8824
3. Left Grip	.9468
4. Sit-Ups (2-minute)	.9630
5. Floor Push-Ups	.9701

Standard Score Equation:

$$\tilde{Z}_0 = .4858Z_1 + .3569Z_2 + .3313Z_3 + .1488Z_4 + .1435Z_5$$

TABLE XXIV

ANALYSIS OF FIVE-ITEM TEST BATTERY NUMBER FOUR UTILIZING THE WHERRY-DOOLITTLE TEST SELECTION METHOD

Test Items in Order of Selection	$\tilde{R}_0$
1. Baseball Throw for Distance	.7452
2. Three Standing Broad Jumps	.8824
3. 100-Yard Pick-a-back	.9633
4. 300-Yard Shuttle Run (150-yard shuttle distance)	.9870
5. Floor Push-Ups	.9921

Standard Score Equation:

$$\tilde{Z}_0 = .8547Z_1 + .8065Z_2 - .7477Z_3 + .2305Z_4 + .1206Z_5$$