

ARMY MEDICAL RESEARCH LABORATORY

FORT KNOX, KENTUCKY

REPORT NO. 3

Reprinted from
Armored Medical Research Laboratory Report 56-1 (07)
19 February 1947

A CRITIQUE OF PHYSICAL FITNESS TESTS*

*Sub-project under Studies of Fatigue in Relation to Military Tasks
(AMRL-56). Approved by CG, ASF, 31 May 1946.



MEDICAL RESEARCH AND DEVELOPMENT BOARD
OFFICE OF THE SURGEON GENERAL
DEPARTMENT OF THE ARMY

REPORT NO. 3

A CRITIQUE OF PHYSICAL FITNESS TESTS*

by

W. B. Bean, Lt. Col. MC¹, C. R. Park, Capt., MC,
D. M. Bell, Capt., RCAMC² and
C. R. Henderson, Capt., SnC.³

from

Armored Medical Research Laboratory
Fort Knox, Kentucky
February 19, 1947

*Sub-project under Studies of Fatigue in Relation to Military Tasks
(AMRL-56). Approved CG, ASF, 31 May 1946.

1. Separated from Service Now at Cincinnati General Hospital,
Cincinnati, Ohio.
2. Collaborator from Canadian Army.
3. Collaborator from Medical Nutrition Laboratory.

ABSTRACT

A CRITIQUE OF PHYSICAL FITNESS TESTS

OBJECT

This project has three purposes: (1) To analyze data from this Laboratory on physical fitness as measured by the Harvard Step Test, the Navy Step Test, the Army Ground Forces Test and the Army Air Forces Test; (2) To discuss the difficulties in definition and measurement of physical fitness; (3) To make recommendations for the improvement of present tests and for the development of new tests.

DISCUSSION

Since military operations require men who are physically fit it is highly desirable that some measurement or test be available to permit discrimination between degrees of fitness. Only by evaluating fitness is it possible to employ preselection logically, to measure the effects of training and to determine stages of convalescence. Since there is no universally accepted definition of physical fitness, many tests designed to evaluate it actually measure different aspects of fitness. When it became apparent that non-performance tests were thoroughly unreliable as predictors of performance, urgency of the war situation did not permit a critical study of the various elements in physical fitness which ought to be measured by an acceptable test. Practicable though empirical methods were employed without a basic study of how well they actually measured the sum total or discriminated between the several component parts of physical fitness. As gross errors became apparent changes were introduced into the tests or scoring systems. Over a period of three years this Laboratory conducted a series of fitness tests under controlled conditions. Since methods and procedures were not changed during this period the data may be used for comparative purposes. Since study of fitness is as pertinent to conditions of peace as to those of war our experience is presented critically in order that future workers may be aware of the complexity and pitfalls of the problem and to suggest lines of future investigation which should clarify the concept of physical fitness. The analysis and discussion do not present a flattering picture of the tests but it is emphasized that they have served an extremely useful purpose during the emergency period. The less urgent times of peace permit a basic and comprehensive reconsideration of the whole problem of testing physical fitness.

CONCLUSIONS

a. None of the tests studied is satisfactory for discriminating between degrees of individual fitness. This fault differs in kind and degree among the tests. It arises from:

1. Failure to test chief components of fitness.
2. Inadequate scoring systems.
3. Abnormal distribution of performance achievement and/or score.
4. Lack of reproducibility.
5. Inability to control or measure motivation.
6. Inequality of stress on all persons.
7. Failure to consider physiologic cost or post-exercise conditions.
8. Presence of test components where readily acquired skills permit subjects to "beat the test".
9. Failure to consider environment or physique in scoring systems.

b. Several of the tests are satisfactory as gross measures of fitness and permit satisfactory comparison of groups.

c. A battery of fitness tests is a better measure than a single test.

d. Appraisal of fitness by good line and non-commissioned officers, familiar with their men, is as good or better than fitness tests in evaluating troops.

e. Performance tests, when competition is aroused, serve as incentives to improve fitness.

RECOMMENDATIONS

a. That a far reaching program of basic investigation in physical fitness and reliable methods for testing it be included in the plan for post-war medical research relating to the army.

b. That the information contained in this report be made available to persons and agencies responsible for physiological research.

c. That until tests are further perfected they be considered as somewhat unreliable aids in evaluating individual fitness, not final determinants.

d. That the tests be considered fairly reliable means for discriminating between degrees of fitness in large groups of men.

Submitted by:

William B. Bean, Lt. Col. , MC
Charles R. Park, Captain, MC
David M. Bell, Captain, RCAMC
Charles R. Henderson, Captain, SnC

Approved /s/ Ray G. Daggs

RAY G. DAGGS

Director of Research

Approved/s/ Frederick J. Knoblauch

Frederick J. Knoblauch

Lt. Colonel, MC

Commanding

REPRINTED

A CRITIQUE OF PHYSICAL FITNESS TESTS

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
Abstract	1
I. Definition	5
II. Comparison of Tests	6
A. Sources of Data	6
B. Harvard Step Test	8
C. Navy Step Test	11
D. AAF Fitness Test	11
E. AGF Fitness Test	14
F. Correlation Among Tests	16
G. Caloric Expenditure	18
III. Environmental Influence on Performance	19
IV. Subjective and Objective Measures	21
V. Relationship of Fitness to Other Factors	22
VI. Historical Review	28
VII. Suggestions for an Improved Test	33
VIII. Methods and Procedures	37
IX. Bibliography	45
X. Figures	

I. A DEFINITION OF PHYSICAL FITNESS

Physical fitness is a term which has been applied to many phases of health and performance. Though its basic importance is widely recognized its definition is vague. To the physician it may signify absence of disease, to the athletic coach the perfection which comes from a program of training and to the employer it may mean satisfactory productivity in labor or industrial work. In terms of military tasks, fitness signifies something special and not interchangeable for the infantryman, the fighter pilot, and the submariner; fitness for attacking a tropical beachhead and an arctic pillbox may not be the same.

Physical fitness as the term is used in this report includes various attributes and is dependent upon the proper interplay of several functions. Physical fitness of whatever kind depends upon (1) a physique or anatomical structure permitting various activities, (2) a physiologic state compatible with carrying out the designated tasks, and (3) will-to-do which directs the person to do the job. In addition skill, a compound of native ability and training, influences performance. A measure of fitness should determine the resultant of these forces at a given time under set circumstances. Its utility hinges on applicability of the measurement to broader fields of performance than reside in the brief small scope of a fitness test.

From the military viewpoint, structural and functional components of physical fitness as well as motivation are requisites for effective performance. A test which would measure them separately would be useful since compensation, by masking a defect in one or another attribute, may reduce the likelihood of potential improvement. Strong motivation even with mediocre structure and physiologic state may yield better performance than poor motivation associated with excellent physique and functional state. Superior physiologic status may compensate for defects in structure. If the will-to-do is poor no test will assess physique and functional state. Therefore, present fitness tests can do no more than appraise the resultant of all factors contributing to fitness. They do not discriminate between or measure separate components. In specific terms physical fitness should include (1) capacity to endure for considerable periods of time multiple types of work on a high plane of energy expenditure, with (2) minimal disturbances of cardiorespiratory, muscular and other physiologic functions and (3) capacity for purposeful activity following work. A test should measure both accomplishment and cost. One must distinguish physical from medical fitness, structural from functional fitness, and soundness (endurance) from momentary fitness.

During the war the need for a simple but reliable test for fitness was urgent. Since no available test gave a satisfactory measure of performance a number of new ones were devised and have been used extensively. It is

recognized that the tests have had manifold usefulness but they also have faults, some of which may be corrected by changing the scoring system or introducing new components or measures into the test. On the basis of the large body of data collected in various tests and surveys conducted by this laboratory, we have analyzed four widely used tests, pointed out defects and suggested methods of improving them.

II. COMPARISON OF TESTS

A. Sources of Data.

1. Fort Knox Studies: A total of 125 men was studied at Fort Knox during the winter and spring of 1943-1944 in order to compare their fitness ratings by the Harvard Fatigue Laboratory Step Test, the Navy Step Test, the Army Ground Forces Test, and the Army Air Force Test. All men were healthy enlisted volunteers between the ages of 18 and 33 years, with average age 21 years. They varied considerably in size and weight and recent physical training. The Navy and Harvard Step Tests were performed in an air-conditioned laboratory on a linoleum composition floor, the AGF and AAF tests were performed outdoors. All tests were run in the morning at least 2 hours after breakfast but the AGF Test was not done on the same day as the others. Rest periods of 45 to 75 minutes separated successive tests (AAF, Navy, and Harvard Step Test) while 15 to 20 minutes separated components of the AGF Test. Smoking was prohibited 15 to 20 minutes before a test. For further details see reference (1).

2. Colorado Studies: A battalion of 827 riflemen was used as subjects in an eight week study. These men, receiving final training for combat, were acting as subjects for the testing of field rations. The Harvard Step Test, the Army Air Force Test, and Army Ground Force Test were conducted at weekly intervals. The measurement of improving fitness under vigorous field activity in unusually well controlled conditions could thus be readily observed.

a. Subjects: Significant data are listed in Table 1.

TABLE 1

Characteristics	Range	Average
Age (years)	18-41	23.7
Weight (pounds)	111-215	152.8
Height (inches)	58-76	68.8
Length of Army Service (months)	6-149	21.9

b. Environment: The tests were conducted in the Pike National Forest in the Rocky Mountain area of central Colorado. It was an isolated area of rugged rock and timbered mountains, rolling hills and valleys and wide plains. The climate was temperate, with the maximum daily temperatures ranging from 72° to 92°F and minimum temperatures from 32° to 45°F. The altitude varied from 8700 to 9000 feet. (All subjects had spent several months at 6100 feet immediately prior to the test period.)

c. General Organization and Activity: The battalion was divided into six (6) companies. Training of all companies was uniform and each week's quantity of work was approximately equal to that of any other week. Insofar as possible intensive infantry combat training, consisting mainly of practical field work was given; lectures were held to a minimum. Training included marches both night and day, combat firing, platoon and squad tactics, organization of the army, outpost problems, map reading and compass work, scouting and patrolling, tactical training of the individual, transition firing, bayonet training, field fortification, foxholes, grenade training, and night vision. Morale of the test subjects throughout the entire period was excellent. A spirit of competition between companies and between platoons within each company was maintained throughout and provided incentive in fitness testing.

d. Organization of Testing: A routine test day involved the following procedures: (a) weighing all men, (b) biochemical studies, (c) a clinical examination, (d) the Harvard Step Test, (e) the Army Air Force Test and (f) the Army Ground Forces Test which was carried out in the afternoon. The battery of fitness tests was given six (6) times. Test 1, in which the AGF Test was not included, was done at 6100 feet altitude; all others in the test area at 9000 feet. Test 2 was done the first full day in the test area. Test 3, done 7 days later, measured effects of acclimatization. The Step Test and the AAF Test were done in the morning, an hour separating the two. Half of the subjects did the Step Test first and half did the AAF Test first. The original sequence was followed by each subject in all subsequent tests. Order of sequence had no apparent effect on the scores. The AGF Test was begun an hour after lunch. Each component was done in the same sequence and interval rest sufficient only to catch the breath was allowed. The 4-mile march did not begin until 30 minutes after the zigzag was done. For further details see AMRL Report on Project No. 30, dated 22 November 1944 (2).

3. Pacific Study: The Harvard Step Test was done on selected subjects on Hawaii, Guadalcanal, Guam, Iwo Jima, and Luzon during the course of a nutrition survey.

The data were taken from samples of at least 50 men who had the characteristics listed in Table 2.

TABLE 2

Location	Age	Overseas	Percent		
			White Troops	Height	Weight
Hawaii	29	23	80	68.7	158
Guadalcanal	28	20	82	68.8	155
Guam	26	21	72	67.9	154
Iwo Jima	26	17	82	69.1	150
Luzon	25	15	100	68.9	144

For further details see Armored Medical Research Laboratory report on Nutrition Survey in Pacific Ocean Areas dated 22 August 1945 (3).

B. Harvard Fatigue Laboratory Step Test: The Harvard Step Test attempts to measure fitness using two criteria (1) the duration up to the 5-minute limit of stepping up and down on a 2-inch platform and (2) the pulse rate for 30 seconds beginning 1 minute after cessation of this effort. To attain good scores the subject must have both good mechanical strength and ample cardiac reserve. Ideally, the measurement of pulse rate in recovery should be made after a standard task, and measurement of muscular strength should be independent. This has been attempted with only partial success in the Navy Step Test. In an attempt to make the procedure as simple as possible the Harvard Step Test combines these two components.

1. Colorado Data: Figure 1 shows the distribution of duration of exercise on the Step Test. On Test 2, 73% of men completed the full 5 minutes of effort and 96% of men on Test 6. In all 2500 tests conducted, 85% of men completed the full 5 minutes.

Distribution of the times achieved by men who failed to complete the full 5 minutes (Fig. 1) shows that very few men stopped between 4 and 5 minutes. The subjects were told how long they had been working and, presumably when within 1 minute of their goal, they expended the extra effort required to continue to the end.

An empirical relationship between performance time and pulse rate govern the scoring system as shown in Figure 2. This system gives 60 points for 5 minutes of effort and the remainder of the score is derived from the pulse rate response. The separation of 85% of soldiers

into more and less fit men thus depends entirely on cardiovascular response to a standard severe task. The pulse rates of these men fall into a symmetrical distribution curve (see Fig. 3), which suggests that scoring for this group should be a linear function of the pulse rate rather than an exponential function as is now the case.

In this study only 15% of men failed to complete 5 minutes of stepping but in a group of less fit men this percentage would be much larger. When less than 5 minutes is completed the actual time of performance greatly influences the final score, and the pulse rate influences it to a lesser extent. In proportion to the mechanical weakness of the subject his score will be reduced. To demonstrate fitness in this group comparable to the group completing the full 5 minutes, there must be a definite correlation between mechanical and cardiovascular strength and it must be properly weighted in the scoring system. The evidence that this is not the case is as follows:

a. The distribution of scores for men completing 5 minutes follows a symmetrical curve (Fig. 4). * The addition of men failing to complete 5 minutes distorts this curve.

b. Satisfactory distribution curves of scores (Fig. 6) was obtained in the Colorado tests when 85% of men completed the full 5 minutes. When a smaller percent completed 5 minutes, the distribution curve was greatly distorted. This can be seen in the curve marked "Hawaii and Guadalcanal" of Figure 6.

c. The heart rate of men completing 5 minutes on the Step Test correlates very poorly with components of the AAF and AGF Tests in which mechanical strength is the chief requirement for a good score. An example is shown in Figure 5.

d. The scores of men who do not complete the full 5 minutes vary much more than the scores of those who do complete the required time. Although this may result from improper motivation or other factors, it introduces an irregularity in the test, particularly in the low score range.

e. Scores made by men completing the full 5 minutes correlate well with AAF Test scores, while the Step Test scores of those failing to complete 5 minutes correlate very poorly.

*A slight distortion of the pulse rate distribution curve from which the score curve is derived is the result of the non-linear relationship between pulse rate and score previously noted. One obvious deficiency is the absence of any scores of 85 which is an artifact arising from the use of the scoring grid (Section VIII, Table 12).

f. Vagal bradycardia may produce a spurious score not really related to fitness under certain conditions (4).

2. Pacific Data: Results of the Harvard Step Test in the Pacific Nutrition Survey are presented in condensed form in Figure 6. Since the distribution curves for Hawaii and Guadalcanal were nearly alike they were combined, as were those for Guam and Iwo Jima. The data from the Pacific have been compared with scores from the first and last test in the Colorado Ration Trials. The distribution curves fall into three distinct groups with low, medium and high scores. The low scores made by subjects on Hawaii and Guadalcanal may be explained only in part by the higher average age and greater weight of the subjects, both of which are associated with lower scores. Though the differences were not large, the environmental factor of heat load was greatest on Guadalcanal and least on Hawaii. The score indicates a low state of fitness consistent with sedentary work and lack of arduous exercise. The distribution curve for the combined data from Guam and Iwo Jima is quite similar to the curve for the first test in the Colorado infantry battalion, although the mean score for the latter is 2 points lower. This is interpreted as indicating a very similar state of fitness in the two groups--a state of average fitness in garrison troops without active training. The highest scores were made by the Colorado test subjects at the end of 8 weeks' intensive training in the field. Distribution of scores from the infantry division in the lines on Luzon is strikingly similar. Age and weight were nearly alike in these groups. It is concluded that the distribution and mean values for Step Test scores of these two groups of subjects indicate a high level of fitness consistent with either effective training or vigorous combat activity and associated with high morale.

Distribution curves in Figure 6 fall into 3 distinct ranges, a poor, an intermediate, and a good. These curves agreed with the observer's impression of the actual state of the men. The test, therefore, has utility in separation of groups, regardless of its defects in evaluating fitness in a single person.

The studies using one simple fitness test demonstrate its utility in field studies when lack of personnel, apparatus and time require a simple rapid test.

Practically, the Harvard Step Test is very easy to carry out, requiring little apparatus. One observer can process 10 or more men an hour. It can be done in the field where more complex tests would be impossible. Subjects dislike the test because of the strain on the leg muscles which often produces soreness, and the dyspnea and fatigue which are out of proportion to the energy used. These objections indicate that the test really taxes the subject.

From these observations it appears that the Harvard Step Test uses two distinct elements of physical fitness--cardiovascular strength and mechanical strength--in a combination which does not permit strict comparison of men within a test group except the very fit men who complete 5 minutes of stepping. Despite these limitations the test is a useful one and serves to give an approximate overall evaluation of the fitness of a group of men

C. Navy Step Test: The inclusion of a distinct cardiovascular part and endurance part in the Navy Step Test is an attempt to include the two chief components of fitness. The distribution of scores skews markedly to the right and is very asymmetrical (5). Because the score is very largely determined by the endurance component, the test loses much of its potential value. In addition, it requires a preceding period of rest, and several observations of pulse rate, rendering its administration to large groups very difficult. Karpovitch has made an analysis of the AAF, Harvard and Navy Tests and found that the test-retest reliability of the Navy Test gave an R value of only +0.48. Studies in this laboratory (1) pointed to the same conclusions independently. Therefore the Navy Test was not included in the battery of tests carried out in the Colorado Ration Trials. Revision of the scoring system would improve the usefulness of the test.

D. The AAF Fitness Test: The 3 components of the AAF Test are a 300-yard shuttle run, sit-ups, and pull-ups. The AAF score is the average of the scores for each component.

1. Three Hundred-Yard Shuttle Run: In the shuttle run the subject must run five 60-yard laps, making a 180° turn at the end of each except the last. The score is based on the time required to traverse the entire course, and a good score requires both sprinting speed and agility in making the turns. The very poor correlations of this test with the Harvard Step Test suggest that the duration of the run is too short for cardiovascular function to be a limiting factor.

The score of the Colorado group on the run was considerably below the "good" rating. Among the reasons for poor performance were: (a) The sandy terrain which was poor for running; (b) regulation army combat boots were worn after Test 1.

The total AAF scores were relatively much lower than the Harvard Step Test or AGF scores where no such hindrances existed, or affected only a fraction of the test components.

Figure 7 shows the distribution of running time in 2 tests in the Colorado study. Definite improvement is noted. Bunching at the low time (high score) portion of the scale appears with improving fitness. This tendency is presumably the result of some factor, perhaps body configuration, which imposes a limit on performance little effected by improving general fitness. The scoring system of the test recognizes this tendency. A given decrement in running time received more score credit when made at the low time end of the score than the same decrement made in the middle or high end of the scale (Fig. 8). This "correction" is in the proper direction but not sufficient to give a symmetrical distribution of scores.

2. Sit-Ups: In this test component, sit-ups must be performed in a prescribed manner, except that some variation in rate is allowed. The score increases with the number of sit-ups up to 114. Beyond 114 sit-ups no further score accrues. The test places a heavy strain on the muscles of the trunk and pelvis and muscle fatigue is the limiting factor in the number of sit-ups that can be performed.

The distribution of the number of sit-ups on Tests 2 and 6 may be seen in Figure 9. Two features of the curve are of interest. First, a group of men was able to complete the full 114 sit-ups necessary to make a perfect score. In Test 2, this was 6% of the total number of men; in Test 6, 20% of all men. In Test 2, more than 90% of men who completed 114 sit-ups were in a single test company. It is possible that this company used a technique which spared them muscular effort and permitted them to attain perfect scores. However, in Test 6, the men who performed 114 sit-ups were evenly distributed throughout all test groups. No break in the rules for performance of the test could be detected to account for this exceptional performance. The second feature of interest is the extension of the distribution curve toward the high number of performances, in contrast to that of the shuttle run which shows bunching as peak performance is approached. In the shuttle runs it was hypothesized that the mechanical structure of the body imposed a limit on performance which checked increase in score though fitness in general was still improving. In the sit-ups the opposite effect, i. e. improvement in score without corresponding increase in fitness, may arise from learning a knack which enables a man to spare himself muscular effort. Again, score does not accurately reflect general physical fitness.

It could be argued that the ability to learn a knack is in itself a measure of physical fitness, but this does not seem to be the case. The AAF scores of men performing 114 sit-ups are contrasted with the men performing between 60-90 push-ups (Table 3). Whereas

men accomplishing 114 sit-ups scored very much higher on the AAF Test, they did not score significantly higher on other tests.

TABLE 3

	Score		
	AAF	AGF	Harvard
"114" Group	62.3	86.8	81.1
"60-90" Group	48.7	83.5	79.7

The scoring system for the sit-ups gives more credit for increments in performance at the low end of the scale than at the high end. (Fig. 10). This partly offsets the skewing of the distribution curve of performance. The correction is not sufficient to give a symmetrical distribution of scores and it does not affect the men attaining perfect scores.

3. Pull-Ups: The pull-up component of the AAF Test is a measure of the muscular strength of the arm and shoulder muscle group. The test is of short duration and the limiting factor in performance is muscular fatigue.

Distribution curves of performance in Test 2 and Test 6 show symmetrical curves with a symmetrical shift of the entire curve with improving performance (Fig. 11). The score should be in linear proportion to the performance and this is almost the case in the official scoring system. At the extremes of performance there is a slight departure from linearity which has only slight effect on the classification of a small percentage of men.

The division of men into thirds of least, average, and most fit depends on a mean difference of slightly more than 4 chin-ups. However, in the series of tests performed over 57 days the men increased only 2 chin-ups, from 7 to 9. This suggests that only marked gross changes in fitness would be detected by this component of the test.

4. AAF Test as a Unit: In each component of the AAF Test some deficiency has been noted. Each deficiency reduces the reliability of the result for a certain percentage of the men. In the case of sit-ups this percentage may be quite large and will have a considerable effect on the final AAF score.

The distribution of total scores on Test 2 and Test 6 of the Colorado study are shown in Figure 12. As would be anticipated from the distribution curves of the separate test components, this curve is also asymmetrical. There is a pronounced shift toward higher scores from Test 2 to Test 6; however, the form of the curve remains about the same. The improvement in fitness in the AAF Test for the group of men as a whole correlates well with the improvement noted by the Harvard Step Test and the AGF Test.

E. AGF Fitness Test: The 6 components of the AGF Test and their percent contribution to the final AGF score are listed in Table 4.

TABLE 4

Name of Test	Character of Test	Contribution to Final Score
4-mile march	Subject carries pack & rifle*	30%
300-yard run	Two 150-yard laps with 180° turn	20%
75-yard pig-a-back-run	Subject carries man of equal weight	20%
Zigzag run	Combines creeping, crawling, broad jumping	10%
Push-ups	Standard calisthenic exercise	10%
Burpees	Standard calisthenic exercise	10%

*Standardized in the Colorado Test to weigh 20-30 pounds.

1. **Four-Mile March:** In this component, the scoring system penalizes the subject for straggling at each mile marker and again for lateness at the finish. If the subject is on time at each mile and finishes in 50 minutes he receives a perfect score.

Performance in the 4-mile march is shown in Figure 13. In Test 2, 40% of men finished on time and Test 6, virtually 100% of men. It is obvious, then, that for the degree of fitness reached by Test 6 the scoring system will not discriminate at all between the more and less fit men of the groups.

2. **Three Hundred-Yard Run, Pig-a-back Run, and Zigzag Run:** The distribution curves of performance in the running components of the AGF Test show a tendency toward bunching of men as fitness improves. (Figs. 14, 15 and 16). As in the AAF shuttle run, this tendency does not necessarily indicate that fitness is reaching a

maximum, but may only indicate that some mechanical factor such as body construction is limiting running speeds. No correction in the AGF scoring system has been attempted for this trend.

3. Push-Ups: The distribution curve of push-ups has a curious form with improving fitness (Fig. 17). This tendency is noted in Tests 4, 5 and 6. It is the result of a maximum score having been arbitrarily placed at 34 push-ups. The men made great efforts to reach 34 but not to continue beyond that figure, as they would receive no further credit. As in the Harvard Step Test, there is a dip in the distribution curve in the region just short of a perfect score which indicates that men who near the mark probably make an extra effort while those who feel they cannot reach perfection quit before exhaustion. In the zigzag runs and pig-a-back runs where most men were finally making perfect scores they had no reliable guide as to their time and did not slow down.

4. Burpees: The distribution of burpees performed (Fig. 18) is a symmetrical curve and shifts symmetrically with improving performance. As in the AAF chin-up test, however, a small difference in the number of performances has a profound effect on the fitness classification.

5. AGF Test as a Unit: The AGF Test has certain features which should make it the most accurate index of fitness for army use. The first is the fact that it employs 6 components. The lack of correlations found in this study between test components indicate that each component measures a different aspect of fitness or that each is highly unreliable. In either instance greater reliability will be achieved by increasing the number of components. The test components are very similar or identical in many cases to the actual activity of the infantry soldier in the field or combat. In other words, a large part of the AGF Test is a direct measurement of practical military performance.

The use of a large number of components has the disadvantage of making the test difficult to administer. About 15 men are required for the rapid testing of any group larger than 25 subjects. Organizing and measuring the test areas is time and labor consuming.

A very bad defect of the AGF Test is the scoring system. The dotted vertical lines on the distribution curves of performance show the levels of performance necessary for a perfect score. Obviously in many components performance is possible beyond the line of maximum score and no additional credit is given by the scoring system. This effect is seen in the distribution of total AGF scores plotted for Test 2 and for Test 6 (Fig. 19). Satisfactory distribution occurs only for the

lower half of scores on Test 2. In Test 6, bunching of the group has occurred to a great extent because a large fraction of men have reached perfect scores in several components. Clearly the fitness of the group as a whole will not be correctly indicated and estimation of individual fitness within the group will be very unsatisfactory. The scoring system should include the highest degree of performance for which data are available and it should be proportional to the performance distribution curve in a manner to give a symmetrical distribution of scores.* (See AAF Test.)

F. Correlation Among Tests and Test Components:

1. Correlations of Tests: Correlation was poor with the Harvard Step Test scores and both the AAF and AGF Test scores. Correlation was fair between the AAF and AGF Test scores (Table 5).

TABLE 5

Tests	Correlation
Harvard vs AAF	.24
Harvard vs AGF	.26
AGF vs AAF	.68

2. Scatter Diagrams were made to establish correlation between certain test components and groups of test components. To avoid errors due to artifacts of the scoring systems, the scatter diagrams were either plots of actual performance, or new scoring systems were used which were directly proportional to performance. Correlation coefficients were not calculated. The diagrams and estimates of correlation are listed in Table 6.

* This correlation cannot be undertaken from this study because the actual performance times on the 4-mile march were not recorded.

TABLE 6

Test Components	Estimation of Correlation
AGF Burpee vs AGF Push-up	Very Poor
AGF 300-Yard vs AGF Pig-a-back	Very Poor
AGF Push-ups vs AAF Chin-ups	Very Poor
AGF Burpee vs AGF Zigzag	Very Poor
AGF 300-Yard Run vs AAF 300-Yard Shuttle Run	Poor
AAF Shuttle Run vs Harvard Step Test	Very Poor

Test Group	Estimation of Correlation
AGF Test without march vs AAF Test	Fair
AGF Burpee + Zigzag + Push-ups vs AGF Pig-a-back + 300-Yard Run	Very Poor
AAF Test vs Harvard + AGF Tests	Fair

3. Improvement in Fitness: The mean scores made on each test have been plotted for successive days (Fig. 20). Although the correlation between individual tests is not good, the degree of mean improvement in fitness indicated by each test is similar. The rate of improvement in fitness appears to lessen in the last days. This may be an artifact of the scoring system arising from the use of maximum scores in many test components. (See discussion of AGF Test.)

G. Caloric Expenditure in Different Parts of the Fitness Tests:

A calculation of the expenditure of Calories on the 10 different exercises of the 3 fitness tests was carried out on selected subjects. The standard open-circuit Douglas bag technic and Haldane analysis were used in the collection of data. These data, calculated as additional cost over and above the average expenditure for very light activity (100 Cals/hr), are given for the usual performance in total work done:

	<u>Calories</u>	<u>Time of Duration or Number of Times Exercise is Completed</u>
Step Test	61	5 minutes
Sit-ups	35	100 sit-ups
Chin-ups	7	10 chin-ups
300-Yard Shuttle Run	21	60-70 seconds
Push-ups	6	20 push-ups
300-Yard Run	22	60-70 seconds
Burpees	10	20 seconds
Pig-a-back	12	20 seconds
Zigzag	14	30 seconds
4-mile Road March, Pack and Equipment (30 pounds)	<u>448</u>	50 minutes
TOTAL	636	

III. ENVIRONMENTAL INFLUENCES ON PERFORMANCE

A. External Factors: Factors in the external environment influence performance in two ways: (1) they may actually alter fitness as in work at high altitudes or in the heat, especially before acclimatization has taken place; (2) they may interfere with carrying out a set task as, for example, running on a muddy or sandy track. Such effects are independent of the state of fitness as determined under a standard environment without extrinsic interference. Nevertheless, influences of this class have often been disregarded in setting up specifications for performance tests, and no scoring procedure has been established which allows for proper weighting of environmental factors of several types. Any test conducted out-of-doors may be disturbed by rain and wind; is influenced by terrain, by firmness of ground, by mud or dust, by stability of equipment or apparatus and sometimes by glare and sunshine. Constricting, ill-fitting or loose clothes and heavy or poorly adjusted shoes interfere notoriously with running, whereas an obstacle course may be negotiated more expeditiously with protective clothing. A general criticism of fitness tests is their lack of regard for the influence of the external environment upon performance. This has prevented exact comparison of tests in groups when environmental influences may have changed.

B. Intrinsic Factors:

1. Physique: Studies of physical fitness have not advanced to the stage where a separation of the various components of performance may be analyzed. One of the important fields for future investigation is the role of body structure in determining performance. It is well known that different body types may be associated with superior performance in different fields. Thus, the good sprinter or distance runner is apt to have a slim wiry build whereas a wrestler is usually heavier and more muscular. Fitness for one task does not imply fitness for another. Obesity is a concomitant of poor condition but height-weight tables do not differentiate mere fatness from the sounder heaviness which may be associated with excellent physical fitness. Behnke *et al* (6, 7) have shown that specific gravity is a better criterion than poundage since it separates the obese from the muscular. Height and limb length influence performance for purely mechanical reasons. Anthropologic type may affect fitness in a specific fashion, although if an influence other than the suspected role of physique exists it has not been measured. One may believe that racial characteristics, separate from physique, may affect muscular efficiency or other aspects of performance in view of the work done by coolies and groups of laborers. But here, too, the possible effects of training and practice remain to be evaluated against the scarcely measured forces of survival of the fittest in its Darwinian sense. Performance is in part influenced by the course of growth and aging but

whether this is mostly a phenomenon of structural change, of biochemical development or of skill and practice is not known. Similarly, the decay of performance with aging is not resolved into its component mechanisms. If such factors are not evaluated a fitness test may measure structure much or little depending on the type of test.

2. Physiologic State: The physiologic and biochemical determinants of fitness are governed by external as well as inherent forces only a few of which are understood. Proper nutrition is a basic requirement of performance. Many types of nutritional aberrations cause a deterioration of performance. These run the gamut from a bone change resulting from chronic calcium depletion to the effect of acute caloric starvation. The effect of deficiencies in B-complex vitamins upon performance has been studied only recently for a few factors. Water and electrolyte equilibria must be maintained in proper balance for the best fitness. The effect of drugs such as alcohol and analeptics must be evaluated. Muscular efficiency and oxidation processes have received extensive study and have had a marked influence in devising tests to appraise fitness.

C. Miscellaneous: Many additional influences have great importance in performance. Of these the chief is the intangible motivation, morale or will-to-do. It dominates performance and is therefore an integral part of fitness. Without it no test of fitness gives a measure of more than an unknown fraction of potential performance. Additional factors such as time of day, elapsed time since meals, quantity and type of food eaten, sequence of tests if several are carried out in rapid succession, rest, sleep and fatigue all add their effects to the underlying attributes which govern performance. The role of innate coordination, learning to accomplish muscular work with least effort and tricks which reduce energy expenditure in a set task, must be evaluated against the real improvement in fitness which comes from the repeated practice which constitutes training. When environmental conditions such as heat and high altitude are encountered improvement from acclimatization must be separated from genuine enhancement of fitness.

Unless all factors are evaluated and separated insofar as possible, any test of fitness may give spurious answers because of the multitude of environmental conditions which affect performance even when fitness itself remains static. Every possible control must be used to regulate the conditions of a test in order that a score will have significance in meaningful terms. Whenever external influences cannot be eliminated they must be measured and recorded in order to appraise their effect upon the results of any test.

IV. SUBJECTIVE AND OBJECTIVE MEASURES OF FITNESS

The final standard against which physical fitness tests must be judged is actual performance. In order to compare a test score with performance, the latter must have some measure in quantitative units by which a score may be validated or invalidated. In the absence of preselection, job analysis or other objective methods of assignment of personnel on the basis of capacity, the infantry soldier is rated by his line and noncommissioned officers. His duties in the field are allotted on the basis of his superior's judgment. Although this is no infallible criterion it has worked out surprisingly well in the hands of capable leaders. It is the method by which the infantryman is given designated tasks. It is of considerable interest to compare the sum of scores on the 3 fitness tests at Colorado with arbitrary ratings of poor, fair, and good given the subjects on the day of testing by their line and noncommissioned officers. Figure 21 illustrates the results of plotting the sum of the scores on the 3 fitness tests against the percentage of ratings of good, fair, and poor in class intervals of 10. Each of the three ratings forms a clearly defined curve with a location which agrees with what would be expected. Though the ratings are arbitrary and varied somewhat with different officers, the pooled data show a very striking agreement between scores and ratings. A similar procedure carried out by the observers gave no significant correlation between ratings and scores, an indication that the exercise of command and living with their men probably enabled the line officers to form a more just estimate of fitness than mere association did in the case of the observers who had no experience in command.

V. RELATIONSHIP OF FITNESS TO OTHER FACTORS

A. Relationship Between Clinical Signs of Nutritional Significance and Scores on Fitness Tests

Two criteria in evaluating health are clinical signs of malnutrition and performance in tests of physical fitness. Little information has existed upon correlation between signs of nutritional deficiency and performance among either the grossly malnourished or well nourished. In the Colorado Test clinical examination and physical fitness tests were given 4 times on the same day at 2 or 3-week intervals to 6 infantry companies. This afforded an opportunity to see whether performance on fitness tests and clinical signs were related.

Data on 4 complete sets of clinical examinations and fitness tests were assembled on a total of 441 men (1764 examinations and tests). Physical fitness scores for each company-date group were separated according to the presence or absence of each abnormality. The mean differences between normal and abnormal groups were calculated for each fitness test. The 24 "within company-date" mean differences were averaged by weighting each of the harmonic mean of the number of men with the number without the abnormality in that "company-date" subclass. Finally the weighted average "within company-date" differences between normal and abnormal men were tested for statistical significance using the standard deviations.

Of 72 possible correlations, 14 were found to be of statistical significance; of these, 12 were in favor of the normal men. Four considerations render these differences of no practical importance: (1) the differences were all small, rarely amounting to a difference of 5 points whereas the experimental error of the fitness tests is actually larger than this; (2) variations among the clinical observers could easily have accounted for many differences between the so called "normal" and "abnormal" subject; (3) the number with positive physical findings was much smaller than the number without and, in fact, hardly significant; (4) many physical signs were isolated phenomena and not related to deficiency disease syndromes. It appears, therefore, that the small but statistically significant differences between certain clinical abnormalities and performance scores are actually of no practical importance. Men rated lowest clinically made practically as good scores as those rated highest; men with best performance and practically as high an incidence of clinical abnormalities as those with worst performance. It appears that, in a normal group, small aberrations in clinical signs are inconsequential in terms of fitness scores.

B. Relationship Between Biochemical Levels in Blood and Urine, and Performance on Physical Fitness Tests

One of the requisites for good performance is a proper function of the physiologic and biochemical systems which govern muscular and cardiovascular fitness. Little is known of the relationship between performance and the vitamin content of blood and urine in a large group of healthy young men. Biochemical determination on hemoglobin, serum protein, serum and urine chloride, fasting and load ascorbic acid, thiamine, riboflavin and F_2 factor in the urine, were made on the same day as the fitness tests. Data for all men with 4 complete sets of observations were calculated for "within company-date" correlations between scores on each fitness test and the 12 chemical determinations. Of the 36 correlations coefficients only 5 were significant, three of these being negative, and all very small. The positive correlations between AAF scores and fasting riboflavin and load ascorbic acid are not considered to have any real meaning. It is concluded that reasonably healthy and fit young men there is no important correlation between vitamin levels and scores on fitness tests.

C. Relationship of Age, Height, Weight for Height and Recent Caloric Intake to Physical Fitness

If fitness tests are of help in evaluating fitness and nutritional status it is essential to know how performance is related to age, height, weight, and recent food intake. Data on age, height, fasting weight, 3 fitness test scores and 10 individual events were available from the ration test material. For 2 of the test periods caloric intake for the preceding 3 weeks was recorded for each subject. For all men with complete data "within company" correlations were calculated and tested for significance for (1) age and fitness test scores, (2) weight and fitness test scores, (3) weight in excess of average for corresponding height and fitness test scores, and (4) caloric consumption for the preceding 3 weeks and fitness test scores.

1. Age: Age was negatively correlated with scores on all 3 tests at each of the 4 periods studied (Tables 7 and 8). In separate events this correlation occurred with AAF Test sit-ups and run, but not pull-ups. In the AGF Test the correlation occurred in the burpee and the shuttle, pig-a-back, and zigzag runs. The regression of scores on age was not linear. There was a tendency for scores to be about the same for ages up through the middle twenties and then to drop off fairly sharply though not very much. The test scores for different age groups are given in Table 8. In the AAF Test on the first day, a rather sharp decline in scores began after the age of 24; this

break came after the age of 27 in the last test. In the first AGF Test, the gradual decline began after the age of 22; in the last test, it began only after the age of 29 and was much less marked. In the first Step Test, the decline came after 26; in the last test, a real decline came only after the age of 32. Insofar as the improvement in score indicates enhanced fitness, it may be said that the effect of age is not noticed in trained men as early as in untrained men. Improvement was nearly the same for all ages in the AAF Test, but the older men (with lower scores) improved more than the younger men in the other tests.

2 Height: Similar correlations were carried out between height and scores on fitness tests (Tables 7 and 9). Height was not correlated with Step Test scores, but was with AAF scores and on the initial test only with AGF scores. Previous studies by Pace (5) indicated a lack of correlation between Step Test scores and height for the Navy Test done on an 18-inch platform. Data from the Harvard Fatigue Laboratory indicate that only extremes of height affected scores by handicapping the very short and facilitating the very tall. In the events of the AAF Test, taller men tended to do more sit-ups and to make faster time on the runs but did fewer pull-ups. The well-known handicap of the short-legged man, and the mechanical disadvantage in height of lift in pull-ups seem satisfactory as an explanation. A priori, one might expect the tall men to encounter more difficulties in the sit-ups owing to the lower arc through which the upper half of the body must bend, but this did not prove to be the case. In the AGF Test, also, the taller men tended to make better times on the shuttle run and more tended to finish the 4-mile march on time. They did fewer push-ups. It appears that mechanical reasons probably account for the differences in performance between tall men and those of average height, although the sit-ups may be an exception.

3. Weight for Height: In each height range, the heavier men tended to make lower Step Test scores, do fewer pull-ups and make lower AAF Test scores. These differences were more pronounced on earlier tests and in some cases had disappeared by the last test. Except for a poorer score on the zig-zag on the first test, there was no correlation between AGF Test scores or events and excess weight for height. Improvement in Step Test and AAF Test scores was directly correlated with loss of body weight.

4. Calorie Consumption: There was a highly significant positive correlation between calorie intake for the first 3 weeks of the test and AAF and AGF Test scores at the end of the period. (Table 10). Men who ate more tended to make faster times on the pig-a-back and shuttle runs, do more burpees and more of them finished the 4-mile march on time. These differences were not so evident by the last test where there had been a general improvement in performance.

TABLE 7
SIMPLE CORRELATIONS OF PHYSICAL FITNESS PERFORMANCES
WITH AGE, HEIGHT, WEIGHT IN EXCESS OF AVERAGE, AND PRECEDING CALORIE INTAKE

Variables	Day of Test	Physical Fitness Test											AGF Score
		Harvard Score	Sit-up	Pull-Up	Run	AAF Score	Push-up	Run	Burpee	Pig-a-Back	Zigzag		
Age	D	-.18*	-.17*	-.06	.20*	-.24*	-.07	.17*	.01	.19*	.14	-.20*	
	D + 21	-.16*	-.16*	-.09	.13	-.22*	-.10	.18*	-.20*	.21*	.10	-.16*	
	D + 35	-.19*	-.17*	-.03	.15	-.20*	-.07	.17*	-.12	.13	.10	-.12	
	D + 56	-.13	-.14	.01	.16*	-.16*	-.11	.16*	-.11	.23*	0	-.18*	
Height	D	.01	.21*	-.28*	-.10	.08	-.31*	-.18*	.01	-.06	-.05	.17	
	D + 21	.03	.24*	-.23*	-.13	.13	-.22*	-.13*	-.10	-.01	.02	.09	
	D + 35	.06	.25*	-.23*	-.15*	.15*	-.24*	-.19*	-.04	-.06	.09	.08	
	D + 56	-.01	.23*	-.23*	-.07	.11	-.22*	-.07	0	-.08	.03	-.05	
Wt in Ex- cess of Av for Ht	D	-.30*	-.07	-.36*	.12	-.20*	.08	.04	.02	-.09	-.12	-.01	
	D + 21	-.14	-.05	-.24*	.01	-.21*	-.04	-.01	-.01	-.07	-.02	.07	
	D + 35	-.12	-.06	-.23*	.01	-.13	.05	.04	.01	-.10	.03	.02	
	D + 56	.01	-.02	-.16*	.05	.05	-.01	.11	0	-.01	.10	-.09	
Cal Intake D to D+20		.11	.07	.09	-.20*	.19*	.11	-.23*	.13	-.16*	-.06	.28*	
Cal Intake D to D+55		.10	0	.06	-.11	.06	.05	-.20*	.01	-.07	-.18*	.12	

* Statistically significant

TABLE 8

PHYSICAL FITNESS SCORES MADE BY DIFFERENT AGE GROUPS

Age Group	Test											
	Harvard				AAF				AGF			
	D	D+21	D+35	D+56	D	D+21	D+35	D+56	D	D+21	D+35	D+56
19-20	66	80	79	84	39	46	47	50	76	86	86	89
21-22	66	78	79	84	38	43	46	49	76	85	86	90
23-24	70	78	84	85	39	42	46	49	76	86	88	89
25-26	66	72	78	82	35	43	45	48	72	85	86	89
27-28	66	75	78	80	37	42	45	47	74	85	86	89
29-30	61	75	76	84	34	41	44	47	72	84	84	87
31-up	57	72	71	77	33	38	41	46	70	83	84	86

TABLE 9

PHYSICAL FITNESS SCORES MADE BY DIFFERENT HEIGHT GROUPS

Height Group	Test											
	Harvard				AAF				AGF			
	D	D+21	D+35	D+56	D	D+21	D+35	D+56	D	D+21	D+35	D+56
61-64	71	79	81	91	37	41	43	48	71	83	83	91
65	64	72	76	81	37	42	46	50	72	85	89	90
66	66	76	75	82	37	43	45	48	73	86	87	89
67	65	79	80	84	37	42	45	47	75	85	85	89
68	62	73	76	79	36	41	44	47	72	84	85	88
69	67	73	80	86	35	40	41	45	74	84	85	88
70	63	81	83	85	37	44	48	52	76	86	85	89
71	63	77	76	80	40	45	49	51	79	88	88	89
72	68	78	80	83	39	46	51	52	77	87	88	90
73&up	71	76	80	83	39	45	48	50	78	86	87	88

TABLE 10

PHYSICAL FITNESS SCORES MADE BY MEN CONSUMING DIFFERENT NUMBERS OF CALORIES DURING THREE WEEKS PRECEDING THE TESTS

Daily Calorie Consumption	Average Physical Fitness Test Scores		
	Step Test	AAF Test	AGF Test
1450 - 2319	75	41	82
2320 - 2449	64	38	80
2450 - 2589	78	38	82
2590 - 2679	76	43	84
3000 - 3139	74	42	85
3140 - 3269	77	42	85
3270 - 3409	80	44	85
3410 - 3539	78	43	86
3540 - 3679	78	42	84
3680 - 3819	77	45	89
3820 - 3949	75	45	90
3950 - 4449	86	48	93

VI. HISTORICAL REVIEW

The fundamental importance of performance is epitomized in evolutionary terms as "survival of the fittest". This was recognized long before subjective estimates or objective measures of fitness were ever systematized. Civilizations based on the work output of slaves or the performance of soldiers understood the practical aspects of physical fitness. Although attempts at precise measurement are modern, thousands of years ago Chinese folk medicine employed a breath holding and pulse counting test for longevity and similar methods are still employed. The Athenian stress on physique and the Spartan stress on ruggedness and endurance emphasized two aspects of fitness which enjoyed a place in the state religions of antiquity. Nevertheless, it has been only in recent times that an objective approach to the problem has been provided by the development of physiology and allied sciences.

Fitness tests have been classified as performance and non-performance (8) and more elaborately into (1) anthropometric, (2) physical performance, (3) respiratory-circulatory, (4) cardiovascular and (5) cardiovascular-physical performance tests (6). Using the latter classification some of the better known tests are considered in this section.

A. Anthropometric: This method of evaluating fitness is based chiefly on stature, sitting height and chest measurements and ratios of weight and height. Although such information has been used to supplement other data, the Army Air Forces (9) have indicated that anthropometry may be used extensively in the selection of pilots. Heath, et al. (10) consider the masculine component in the selection of officer candidates and show its relation to physical fitness as judged by the Harvard Step Test. Further evaluation is needed before reliance is placed too exclusively upon morphology alone.

B. Physical Performance: The first tests to be used as a gauge of general fitness were based chiefly on strength. Weight lifting and dynamometers for testing strength of various muscle groups are still used and are of limited value. Calisthenic exercises have been used extensively. They include the Army Air Forces Test and the Army Ground Forces Test which are described in detail later. The Army Air Forces Test (11, 12) was devised in an attempt to define and measure elements of fitness required for duty with the Air Forces. Seven elements were considered important and a battery of 15 tests was devised to measure them. These were first reduced to 7 and later to 3 tests which had a correlation coefficient of 0.90 with the original 15 tests. In addition to the Ground Forces Test, others used by the Army include obstacle course runs with score based on time, and an endurance hike with full field pack with score based on the time required to complete the hike.

Another test to measure motor fitness is the Illinois Motor Fitness Screen Test (8), composed of 14 components which attempt to measure 6 elements of motor fitness: balance, flexibility, agility, strength, power, and endurance. Additional requirements include swimming ability and rating of physique.

C. Respiratory-Circulatory: During the last war Flack (13) was interested in determining fitness and fatigue in men of the Royal Air Force. He used 6 tests, 5 being based on respiratory function. The 4 most used were: breath-holding test, vital capacity, expiratory force test and persistence test in which the mercury in a manometer was kept at half the height obtained during the expiratory force test for as many seconds as possible without breathing. The behavior of the pulse during this period was noted. The Flack-Woodham Index of fitness of young and adolescent boys was a development directed toward an estimate of physical fitness.

$$F-W \text{ Index of Fitness} = \frac{\text{Pr} \times \text{Per} \times \text{Br}}{100 \times \frac{(\text{Age in Years})}{4} \times 1.807}$$

where Pr = Max. expiratory force in mm. of Hg.

Per = Time in seconds of breath hold in the persistence test.

Br = The time in seconds of the breath holding test.

L. D. Cripps (14) found that variations of the respiratory test even in a highly selected group were so great that fixing a normal standard was impossible.

In 1935, McCurdy and Larson (15) introduced a test in which observations are made on diastolic pressure (sitting), breath holding 20 seconds after a stair climbing exercise, difference between standing pulse and pulse rate 2 minutes after exercise, standing pulse rate and vital capacity. The amount of exercise is determined from a table of age and weight. Scoring is calculated from these tables.

D. Cardiovascular: In 1904 Crampton presented his "Blood Ptosis Test". The scoring of the test was revised in 1913 (16) and 1920. The test is based on the concept that with poor physical condition there is a lack of vasomotor control and vascular tonicity with resulting blood ptosis and a drop in systolic pressure. Good physical condition causes a compensation and the blood pressure rises. Pulse

rates rise in the unfit and remain the same or rise only slightly in the fit. The two elements considered are "an increase in systolic blood pressure which connotes efficiency and an increase in pulse rate which connotes deficiency". Original ranges were found to be +10 to -10 for changes of systolic blood pressure, and 0 to +44 for pulse increase. "Upon a statistical balancing of these two series of frequencies, the assigning equal percentages to equal ranges, a scale was constructed for evaluation". In 1920 this was extended (17) to give values for increases in heart rate as high as 80/min. and systolic blood pressure variations of 50 mm. Hg.

Meylan, (18) in 1913. judged efficiency by the following: (a) weight, color of skin, and general appearance such as firm vigorous muscles, (b) pulse rate in the horizontal and vertical positions, (c) systolic blood pressure in the horizontal and vertical positions, and (d) heart reaction after hopping 100 feet.

Foster, (19) in 1914, introduced a test involving heart rate in the quiet standing position, immediately after running in a fixed place for exactly 15 seconds at a rate of 180 steps per minute, and 45 seconds after cessation of the exercise.

In 1917, Barringer (20) introduced a test based on the "delay rise" of systolic blood pressure following exercise. He believed that a delayed rise represented an overtaxing of the reserve power of the heart and was associated with a prolonged fall toward the normal resting level. Increasing amounts of work were given the subject at widely separated intervals until a "delayed rise" was elicited.

Sewall (21) later showed that a weakened patient may not have a systolic drop as indicated by Crampton, but a rise of diastolic pressure and a small pulse pressure. He employed these as measures of fitness.

Schneider, (22, 23) in 1920 and 1923, introduced a test which has been used extensively to estimate fitness of pilots. He considered that previous cardiovascular tests were not comprehensive enough. He developed a test which weighs data from 6 sets of observations: pulse rate during recumbency, pulse rate increase on standing, exercise pulse rate, and decline in pulse rate following exercise, resting systolic blood pressure, and systolic blood pressure upon standing.

Turner, (24) in 1927, used a test based on the adaptability of the circulation to quiet standing in one position for 15 minutes and changes in position. A graded scale derived from reclining heart

rate, standing heart rate, general course of the heart rate during prolonged standing and the changes in systolic, diastolic, and pulse pressure while standing was employed.

In 1931, McCloy (25) introduced a cardiovascular test involving only the diastolic blood pressure and heart rate in a quiet standing position. The formula for scoring is $(.89 \text{ S. D. P.}) - (\text{S. P. R.}) + 16$. Ratings above zero indicate a satisfactory state of health.

Graybiel and McFarland, (26) in 1941, considered the use of the tilt table in a test scored on the basis of (a) fainting, (b) the maximal fall in systolic blood pressure below that of the reclining level and (c) the minimal pulse pressure while in the tilted position.

In 1943, Starr (27) introduced a modified cardiovascular test based on pulse and blood pressure in recumbent and erect position, using ballistocardiographic data. The average change in heart rate was +18 and change in blood pressure was +5 mm. Hg. From this he developed the following formula:

$$\begin{aligned} a &= \text{mean pressure change} - 5 \\ b &= 8 - \text{pulse rate change} \\ \text{Index} &= a + b \end{aligned}$$

This test has been used to determine when a patient should resume exercise following illness.

E. Cardiovascular and Physical Performance: In these tests the subject is given work severe enough to tax the cardiovascular system. They had their origin in laboratories where work could be measured accurately by the bicycle ergometer or treadmill and cost determined by O_2 consumption, blood lactate and pulse rate.

In 1942, the Harvard Fatigue Laboratory standardized a treadmill test which was later adapted as a pack test (28) for out-of-doors by providing work equal to that of the treadmill test. The subject stepped up on a 16-inch box 30 times a minute for 5 minutes while carrying a pack of approximately $1/3$ his body weight on this back. Hand grips at shoulder level were provided. This was further simplified as the Step Test without pack.

The Navy or Behnke Step Test described in 1943 is similar in type but more complicated because it is divided into 2 parts and requires several pulse counts.

Specific gravity has been employed by Behnke (6, 7) to separate fit and unfit men especially when they are heavy. Technical difficulties precluded its wide use at present.

Rifle firing has been tried as a measure of performance of infantrymen (1) but has several faults as an objective test.

VII. SUGGESTIONS FOR AN IMPROVED TEST

In Table 11 the 4 tests discussed in this report have been evaluated for each of the factors listed as important in an improved test. Arbitrary ratings range from satisfactory to absent. No test comes near fulfilling all the qualifications of an ideal test.

A. Neither step test taxes many components of fitness. The AAF Test taxes a number while the AGF Test taxes a large number of the components of fitness. It appears that more components can be tested only by multiplying the complexity of actual number of separate parts of a test.

B. Although both the Harvard and the Navy Step Test and the AAF Test involve a fairly high energy output they do so only for certain aspects of muscular exercise and thus cannot tax all components on a high energy level. The Step Tests evaluate high energy output for a few minutes only and the AAF Test does not really tax the performer. Even the AGF Test is unsatisfactory because several of its components do not require a high energy output.

C. The 5-minute limit of the Harvard Step Test can be completed by about 85% of men in good physical condition and beyond this dividing line further separation is lost as far as endurance is concerned. The Navy Step Test has a separate endurance phase though the scoring system reduces its value. The AAF Test does not measure endurance except over very short periods in the pull-ups and chins. In the AGF Test, endurance is measured fairly well by the 4-mile march after the 5 earlier test components.

D. Similarity of stress cannot be achieved where size, shape and aptitude influence performance; therefore, any test in which these factors are important loses some of its accuracy. Since, however, certain aspects of physique may be considered as elements of fitness, a test which may be influenced disproportionately by a physical characteristic fails to differentiate physique from physiologic status. A high score may be obtained by a tall, moderately fit man or a short, very fat one. Tasks which require special skills or coordination have a reduced value in any study where tests are repeated, for a learning curve may obscure true improvement. Thus, in evaluating fitness from test scores, it is important to know whether any peculiar physical trait exists. Scoring systems could introduce a correction for the effects of physique upon scores. Reasonably similar stress occurs in men walking, running and performing customary tasks. Thus the step tests are based on a somewhat artificial situation whereas at least some of the components of the other tests are significantly affected by size.

TABLE 11

An Ideal Test of Physical Fitness Should:	Harvard Step Test	Navy Step Test	AAF Test	AGF Test	Hypothetical Treadmill Test
1. Test chief components of fitness	x	x	Fair		Fair
2. Tax each component on high energy level	x	x	x	Poor	-
3. Measure endurance	Poor	Fair	x		
4. Put reasonably similar stress on all	Fair	Fair			
5. Show little environmental effect	Fair	Fair	Poor	Poor	
6. Consider physiologic cost	Fair	Fair	x	x	
7. Consider post-exercise condition	x	x	x	x	
8. Be independent of motivation	x	x	x	x	
9. Be reproducible		Poor	Fair	Fair	
10. Be simple to conduct			Poor	x	x
11. Be simple to evaluate	x	x	x	x	
12. Have normal distribution of scores		x	x	x	-
13. Have improving scores with improving fitness					
14. Have small learning component	Fair	Fair	x	Fair	

E. The effects of environment on fitness tests have not been studied systematically. It was found that an increase in altitude from 6000 feet to 9000 feet produced striking decreases in the Harvard Step Test and AAF Test scores. Such decreases in score gave only a poor indication of the distress produced by exercise and the relatively poor post-exercise condition of the subjects. Of course heat, rain, terrain and clothing may all exert a profound but as yet not measured effect upon performance. Unless further investigation results in use of factors of correction in the scores the test may be rendered unsatisfactory because of meteorological and environmental changes outside the control of the investigator which are not provided for by the scoring systems.

F. Physiologic cost is not even considered in the AAF and AGF Tests. It is considered only in terms of pulse rate in the step tests, but even this limited observation greatly increases the value of the test. It has been noted that performance in terms of endurance has most weight in the final score. The utility of blood pressure measurements is probably limited but an investigation of respiration and ventilation, even if only a count of respiratory rate, should be investigated.

G. No test takes into consideration the state of the subject after the test though it is obvious that a man who completes a task and collapses is not as fit as one who does the same task and remains in good condition.

H. No test is independent of motivation. In some tests it may actually dominate performance.

I. If a fitness test is not reproducible within reasonable limits it has little value in helping to judge fitness. Errors in procedure, faults of the scoring system, the presence of a large learning component in performance, acquired skill or ability to "beat the test" and variations of environmental factors influence work and efficiency. It is not rare that mere reproducibility signifies a fault in the scoring system as in the 4-mile march of the final AGF Test where more than 99% of the subjects finished on time although there was a wide scatter of times. No separation was made of these men although obviously there were readily appreciated differences among them. It may be argued with propriety that lack of reproducibility may simply indicate true change in fitness. In the absence of any final criterion of evaluation of fitness and lack of a quantitative measure of fitness in the aggregate it remains a matter of judgment as to whether varying scores indicate a fault of the test or a change in fitness.

J. In order that large numbers of men may be processed as rapidly and easily as possible, simplicity is one of the chief goals in fitness testing.

It becomes a question of where oversimplification destroys the significance of a test. Since there is no final standard against which to judge, this can be decided only by the subjective evaluation of fitness.

Other factors remaining constant, the more elements there are in a test or battery of tests, the more likely it is to be a significant measure of true fitness. Information at hand does not allow a decision as to the precise point on the scale from simplicity to complexity where the most information can be gained for the least effort. There is more danger from oversimplification than from over complexity and the reductio ab absurdum of trying to learn almost everything by doing almost nothing is approached in some tests.

K. None of the tests is simple to evaluate because there is poor mutual intercorrelation and there is no quantitative measure of performance against which to evaluate each one singly. In the Colorado Test there was a fairly good correlation between the sum of the scores on the three tests and the company officers' ratings of fitness.

L. Only the Harvard Step Test approaches a binomial or normal distribution of scores; the others all show asymmetry, skewing or bimodality. This is frequently a fault of the scoring system rather than the test itself but in sit-ups and push-ups the limit of improvement in score before fitness has reached a peak partly defeats the purpose of the tests.

M. All tests seem to have improving scores with improving fitness though whether this is a parallel change cannot be stated.

N. The learning component is presumably small in any exercise which is usual in everyday life. Thus walking or running require little if any learning while sit-ups, pull-ups, chins, and burpees are calisthenic exercises in which learning may effect score improvement regardless of changes in actual condition. A learning phase in the sit-up test weakens its value considerably. The ingenuity used to "beat the score" and at the same time avoid extra effort is important but can hardly be measured.

O. A hypothetical treadmill test could be devised to satisfy most of the desiderata except for simplicity in apparatus and conduct of the test.

VIII. TEST METHODS AND SCORING PROCEDURES

The methods actually used in administering the various tests are given in detail because slight variations may affect the score. None of the tests is definitive and changes in directions and scoring systems are still being made by the proponents of some tests.

A. Harvard Fatigue Laboratory Step Test:

1. Stepping boxes 20 inches in height were prepared. The subjects lined up in front of the boxes, stripped to their underwear and socks or bare feet.

2. A pendulum, consisting of a weight on a string 39 inches in length, hung from an improvised scaffold, indicated the required rhythm.

3. At the signal "start" the subject placed one foot on the box, stepped up placing the other foot on the box, straightened the legs and back, and immediately stepped down. At exactly 2-second intervals, the signal, "Up!" was given by the observer. The rhythm was maintained by giving the count "Up-2-3-4, Up-2-3-4". Some subjects responded better to a tap on the back or arm at the required "stepping up" time, while others maintained satisfactory cadence by watching the pendulum. The same foot was used to initiate stepping up and stepping down. The subject was instructed to "lead off" with the same foot each time, although one or two changes during the test were permitted. The swinging of the arms was allowed, but the pressing of the hands against the thighs was forbidden.

4. The "time" began when the subject started exercising. If the subject fell behind the rhythm for 20 seconds without it being regained, he was stopped. No men were allowed to continue for more than 5 minutes. Time was recorded by a stop watch.

5. Upon the termination of exercise the subject was immediately seated and time was counted.

6. The pulse rate was counted from 1 minute to 1 minute 30 seconds following completion of exercise.

7. The duration of effort and the number of heart beats during the 30-second interval were recorded.

8. The score was read from a chart. (Table 12).

TABLE 12
SCORING OF STEP TEST

Duration of Effort	Heart Beats from 1 Min. to 1 1/2 Min. in Recovery										
	40- 44	45- 90	50- 54	55- 59	60- 64	65- 69	70- 74	75- 79	80- 84	85- 89	90 over
0' - 29"	5	5	5	5	5	5	5	5	5	5	5
0' 30" - 0' 59"	20	15	15	15	15	10	10	10	10	10	10
1' 00" - 1' 29"	30	30	25	25	20	20	20	20	15	15	15
1' 30" - 1' 59"	45	40	40	35	30	30	25	25	25	20	20
2' 0" - 2' 29"	50	50	45	45	40	35	35	30	30	30	25
2' 30" - 2' 59"	70	65	60	55	50	45	40	40	35	35	35
3' 0" - 3' 29"	85	75	70	60	55	55	50	45	45	40	40
3' 30" - 3' 59"	100	85	80	70	65	60	55	55	50	45	45
4' 0" - 4' 29"	110	100	90	80	75	70	65	60	55	55	50
4' 30" - 4' 59"	125	110	100	90	85	75	70	65	60	55	50
5'	130	115	105	95	90	80	75	70	65	65	60

Find appropriate line for duration of effort; then find the appropriate column for pulse count; read off the score where the line and column intersect.

- Below 50 - Poor general physical fitness
- 50 - 80 - Average general physical fitness
- Above 80 - Good general physical fitness

B. Army Air Forces' Test: The AAF Test is composed of 3 elements: The sit-up, the pull-up or chin and the shuttle-run. The test subjects wore regulation field uniform and combat shoes throughout the entire test. The jacket was kept on if a 2-piece uniform was used.

1. Sit-Up: The subject began the test lying supine on the ground with hands placed behind head. He sat up, then extended his arms to touch toes with hands, keeping his knees straight and then resumed supine position. No counterweight was used. The subject was not allowed to "bounce" himself up. He kept his hands behind his head until erect and did not rest between sit-ups. Sit-ups were repeated as frequently as possible, but not more than 114 times. The number of complete sit-ups was recorded.

2. Pull-Up or Chin-Up: The subject grasped the bar with the palms facing inward and hung free with the arms fully extended. He then began the exercise by pulling himself down so that arms were fully extended. This was repeated as many times as possible. No kicking or swinging was permitted. The number of complete pull-ups was recorded. There was no time limit. Incomplete pull-ups were not counted.

3. Shuttle Run (300 yards): Two poles were set up in level ground 60 yards apart, the timer at one pole, the subject at the other. At the starting signal, the timer started his watch or noted the time if no stop watch was available, and the subject started his run. The poles were rounded but not touched. Five lengths of 60 yards constituted the test run. The time in seconds was recorded, fractions of seconds being converted to the next full second.

4. Scoring: The score is computed from Table 13.

C. Army Ground Forces' Test: This test is a battery of 6 different tests: the push-up, the 300-yard run, the burpee, the 75-yard pig-a-back, the 70-yard zigzag and the 4-mile march.

Subjects went from one event to another without pause until the 4-mile march, before which they had a half hour rest. Events were run in the order listed. Men wore field uniforms and combat boots throughout the entire test. During the 4-mile march men carried field equipment weighing 30 pounds. (See Table 14 for scoring.)

1. Push-Up: From the leaning rest position, the arms were bent at the elbow until chin and chest were near the ground with the body rigid. The body was raised by straightening the arms. The exercise

TABLE 13
SCORING OF ARMY AIR FORCES TEST

Sit-Ups		Pull-Ups		Shuttle-Run		Sum of Scores	Final Fitness Rating
No.	Score	No.	Score	No.	Score		
114	100	23	100	35	100	300	100
108	98	22	99	36	95	290	98
102	96	21	97	37	90	280	96
96	94	20	94	38	88	270	95
90	92	19	90	39	85	260	93
84	88	18	86	40	83	250	90
78	83	17	82	41	80	240	85
72	78	16	78	42	78	235	81
70	75			43	75	230	78
69	74					225	75
66	73	15	74	44	74	224	74
63	72	14	70	45	72	220	73
60	71			46	70	215	72
57	70	13	66	47	67	210	70
54	68			48	64	205	68
51	66	12	62			200	66
50	64					195	65
						190	64
49	63			49	63	189	63
48	61	11	58	50	62	185	61
45	58			51	60	180	60
42	55	10	54	52	58	175	58
39	52			53	55	170	57
36	51	9	50	54	52	165	55
33	49			55	50	160	54
31	47	8	47	56	47	155	52
						150	50
						145	48
						140	47
30	46	7	44	57	46	139	46
				58	44	135	45
27	43	6	41	59	42	130	44
				60	40	125	42
24	40	5	38	61	38	120	40
				62	36	115	38
21	37	4	35	63	34	110	36
						105	35
20	34					100	34
19	33			64	33	99	33
18	30	3	32	65	25	90	30
15	27			66	22	80	27
12	25	2	29	67	20	70	23
9	22			68	18	60	20
6	13	1	26	69	15	50	17
3	5			70	13	45	15
				71	10	40	10

Instructions: The appropriate numbers are totaled and the final fitness rating is read from the last column. In the number of sit-ups where there may not be a corresponding number on the score table, take it to the closest number.

TABLE 14

SCORING OF THE ARMY GROUND FORCES TEST

Event	Scoring	Weighting Factor
1. Push-Ups	3% for each push-up	1
2. 300-Yard Run	45 seconds or under, score 100%. Deduct 4% for each second (or fraction) over 45 seconds.	2
3. Burpee	9% for each complete burpee	1
4. 75-Yard Pig-a-back	20 seconds or under, score 100%. Deduct 4% for each second (or fraction) over 20 seconds.	2
5. 70-Yard Zigzag	30 seconds or under, score 100%. Deduct 4% for each second (or fraction) over 30 seconds.	1
6. 4-Mile March	For straggling during 1st mile, deduct 8%; during 2nd mile 6%; 3rd mile 4%; 4th mile 2%. At finish deduct 5% for each minute (or fraction) over 50 minutes up to 5 minutes. Failing to finish score, zero. Penalties for any straggling are additive and are added to penalty for failure to finish on time. Straggling shall be construed as more than 1 minute late at each mile marker except at finish where men must be on time.	3

The score achieved on each event is multiplied by its weighting factor to give the weighted score for event. The weighted scores are added, divided by 10 (the sum of the weighting factors) to give the final score for the Army Ground Forces Test.

Assessment of Fitness, rating from final score:

Below 70	Unsatisfactory
70 - 77	Satisfactory
78 - 87	Very satisfactory
88 - 94	Excellent
94 or over	Superior

was repeated as many times as possible. There was no cadence or time limit. Push-ups accomplished by bending or rocking body were not counted. The number of push-ups was recorded.

2. Three-Hundred-Yard Run: The run was 150 yards around a marker and return to the starting line. Time was recorded in seconds, raising fractions of seconds to the next full second.

3. Burpee: From position of attention subject bent to squatting position. The hands were placed on ground inside knees and at the same time legs were extended straight to the rear, the squatting position was resumed and then the position of attention. The exercise was repeated as many times as possible in 20 seconds. The number of complete burpees was recorded.

4. Seventy-Five-Yard Pig-a-Back: Subjects carried men of approximately their own weight. Men who fell down were allowed to repeat. Time in seconds was recorded, raising fractions to next full second.

5. Seventy-Yard Zigzag: Subjects ran 10 yards, crawled 10 yards, ran 10 yards, crept 10 yards, ran 10 yards, jumped 10 yards and ran 10 yards. At the end of each run, except the last two, the subject "hit the ground". The jumps were five feet from center to center of the islands which were 2 feet in diameter. Six jumps, landing on both feet and keeping feet together, were required to cross the 10-yard interval. Direction of course changed 45 degrees each 10 yards, alternating right and left turns. Subjects did not dive when "hitting the ground" but crawled and crept the full 10 yards. Time was recorded in seconds, raising fractions to next full second.

6. Four-Mile March: As each group completed the first 5 components, it assembled with full field equipment and marched over a 4-mile measured course. Times were recorded for each mile of the courses as well as the total time, if more than 50 minutes.

D. Navy Step Test or Behnke Test: This test consists of two elements: a short period of exercise to elicit pulse rate response and a sustained period of moderate exercise to measure muscular endurance. Equipment comprised a box exactly 18 inches in height, a stop watch and a pendulum. The subject wore shorts or underwear, without shoes.

1. Heart Rate Response to Moderate Exercise:

a. The sitting pulse rate was taken after the subject had been seated quietly for 2 or 3 minutes, at least twice to be certain that it was approximately stable.

b. The subject then stood and placed one foot on the step, maintained it there throughout the test.

c. On a signal from the observer, the subject stepped up and down in time with the pendulum 20 times in 30 seconds. The subject stepped precisely with the signal and straightened the knee of the lifted leg as the other foot was placed firmly on the box. At the completion of 20 step-ups the subject sat down.

d. The pulse from 5 seconds to 20 seconds after completion of exercise was converted to rate per minute. The pulse was again recorded from 2 minutes, to 2 minutes 15 seconds following exercise, and converted to rate per minute.

2. Endurance Time and Post-Exercise Heart Rates.

The endurance run began 15 seconds after last pulse reading or 2 minutes 30 seconds after previous exercise. The subject continued the standard exercise until a sharp break in rhythm or exhaustion occurred. Time was recorded to the nearest second.

3. Scoring. The score is determined in accordance with directions given in Table 15.

TABLE 15

SCORING THE NAVY OR BEHNKE TEST

The test is evaluated in terms of two components, the cardiovascular score and the endurance time.

The cardiovascular score is calculated by means of the following equation:

$$C. S. = (B - 70) + 3 (C - A),$$

where A = sitting pulse rate per minute

B = pulse rate per minute immediately after exercise
(5 sec. to 20 sec.)

and C = recovery pulse per minute (120 sec. to 135 sec.
count).

Also, when A is 70 or less, it is considered to be 70 and
when (C - A) is 4 or less, the expression 3 (C - A)
is considered to be 0.

Interpretation of the result and values:

<u>Cardiovascular Score</u>	<u>Rating</u>
Above 74	Poor
51 - 74	Fair
0 - 51	Good

The endurance time is considered to be directly proportional to the physical training of the individual. Interpretation of the endurance time values:

Below 60 Sec.	Poor
60 - 120 Sec.	Fair
Above 120 Sec.	Good

The physical condition as evaluated by these tests is expressed as an index:

$$\text{Step Index} = \frac{\text{Endurance Time}}{\text{Cardiovascular Score}} \times 10$$

In this equation, if the cardiovascular score is 50 or less, it is considered to be 50. An arbitrary set of standards for rating fitness is given below:

<u>Step Index</u>	<u>Rating</u>
Below 8	Poor
8 - 12	Fair
Above 24	Good

IX. BIBLIOGRAPHY

1. Eichna, L. W., Bean, W. B., and Ashe, W. F.: Development of tests to evaluate the fitness of men. Project 5-27, Armored Medical Research Laboratory, Fort Knox, Kentucky, Mar. 10, 1944.
2. Bean, W. B. et al.: Field test of acceptability and adequacy of U. S. Army C, K, 10-in-1 and Canadian Army mess tin rations. Project 30, Armored Medical Research Laboratory, Fort Knox, Kentucky, Nov. 22, 1944.
3. Bean, W. B., Henderson, C. R., Johnson, R. E., and Richardson, L. M.: Report of nutrition survey in the Pacific Theater of Operations, April-June 1945. Nutrition Division of the Office of the Surgeon General, Washington, D. C., Aug. 22, 1945.
4. McPhee, H. R., and Wills, P. V.: Spurious fitness by the endurance test. Journal-Lancet 65: 226, 1945.
5. Behnke, A. R., Welham, W. C., White, W. A. Jr., and Pace, N.: The step-up test to evaluate fitness for physical exertion in healthy men. Project X-134, Report 2, Naval Medical Research Institute, Bethesda, Maryland, Dec. 6, 1943.
6. Behnke, A. R., Feen, B. G., and Welham, W. C.: The specific gravity of healthy men. J. A. M. A. 118: 495, 1942.
7. Welham, W. C., and Behnke, A. R.: J. A. M. A. 118: 498, 1942.
8. Cureton, T. K.: The unfitness of young men in motor fitness. J. A. M. A. 123: 69, 1943.
9. Larsen, L.: Notes on the second fitness conference. Section V, Report to the O. S. R. D., Washington, D. C., Apr. 13, 1943.
10. Heath, C. W., Woods, W. L., Brouha, L., Seltzer, C. C., and Bock, A. V.: Personnel selection: a short method for selection of combat officers. Ann. Int. Med. 19: 415, 1943.
11. U. S. Army Air Forces: Form No. 28A.
12. U. S. Army Air Forces: Regulation 50-10.
13. Flack, M.: Cited by Cripps (14).

14. Cripps, L.D.: The application of the air force physical efficiency tests to men and women. Special Report Serv. 84, Medical Research Council, London, 1924.
15. McCurdy, J.H., and Larsen, L.A.: Measurements of organic efficiency for the prediction of physical condition. Research Quart. Am. Phys. Educ. Supp. 6: 11, 1935.
16. Crampton, C.W.: Blood ptosis, a test of vasomotor efficiency. New York State J. Med. 98: 916, 1913.
17. Crampton, C. W.: The gravity resisting ability of the circulation; its measurement and significance (blood ptosis). Am. J. M. Sc. 160: 721, 1920.
18. Meylan, G.L.: Twenty years' progress in tests of efficiency. Am. Phys. Educ. Rev. 18: 441, 1913.
19. Foster, W.L.: A test of physical efficiency. Am. Phys. Educ. Rev. 19: 632, 1914.
20. Barringer, T.B. Jr.: Studies of the heart's functional capacity. Arch. Int. Med. 20: 329, 1917.
21. Sewall, H.: On the clinical significance of postural changes in the blood pressures, and the secondary waves of arterial blood pressure. Am. J. M. Sc. 158: 786, 1919.
22. Schneider, E.C.: Physiology of muscular activity. Saunders, Philadelphia, 1941.
23. Schneider, E.C.: Cardiovascular rating as a measure of physical fatigue and efficiency. J.A.M.A. 74: 1507, 1930.
24. Turner, A.H.: Adjustment of heart rate and arterial pressure in healthy young women during prolonged standing. Am. J. Physiol. 81: 197, 1927.
25. McCloy, C.H.: Cardiovascular rating of present condition. Arbeitsphysiol. 4: 97, 1931.
26. Graybiel, A. and McFarland, R.A.: Use of the tilt table in aviation medicine. J. Aviation Med. 12: 3, 1941.

27. Starr, Isaac: A simple test of fitness of the circulation. National Research Council, Division of Medical Sciences, O. S. R. D. , Report 133, University of Pennsylvania, May 1, 1933.
28. Johnson, R. E. , Brouha, L. , and Darling, R. C. : Test of physical fitness for strenuous exertion. Rev. canad. de biol. 1: 491, 1942.

X. FIGURES

FIG. 1

DISTRIBUTION OF PERFORMANCE TIME
ON HARVARD STEP TEST

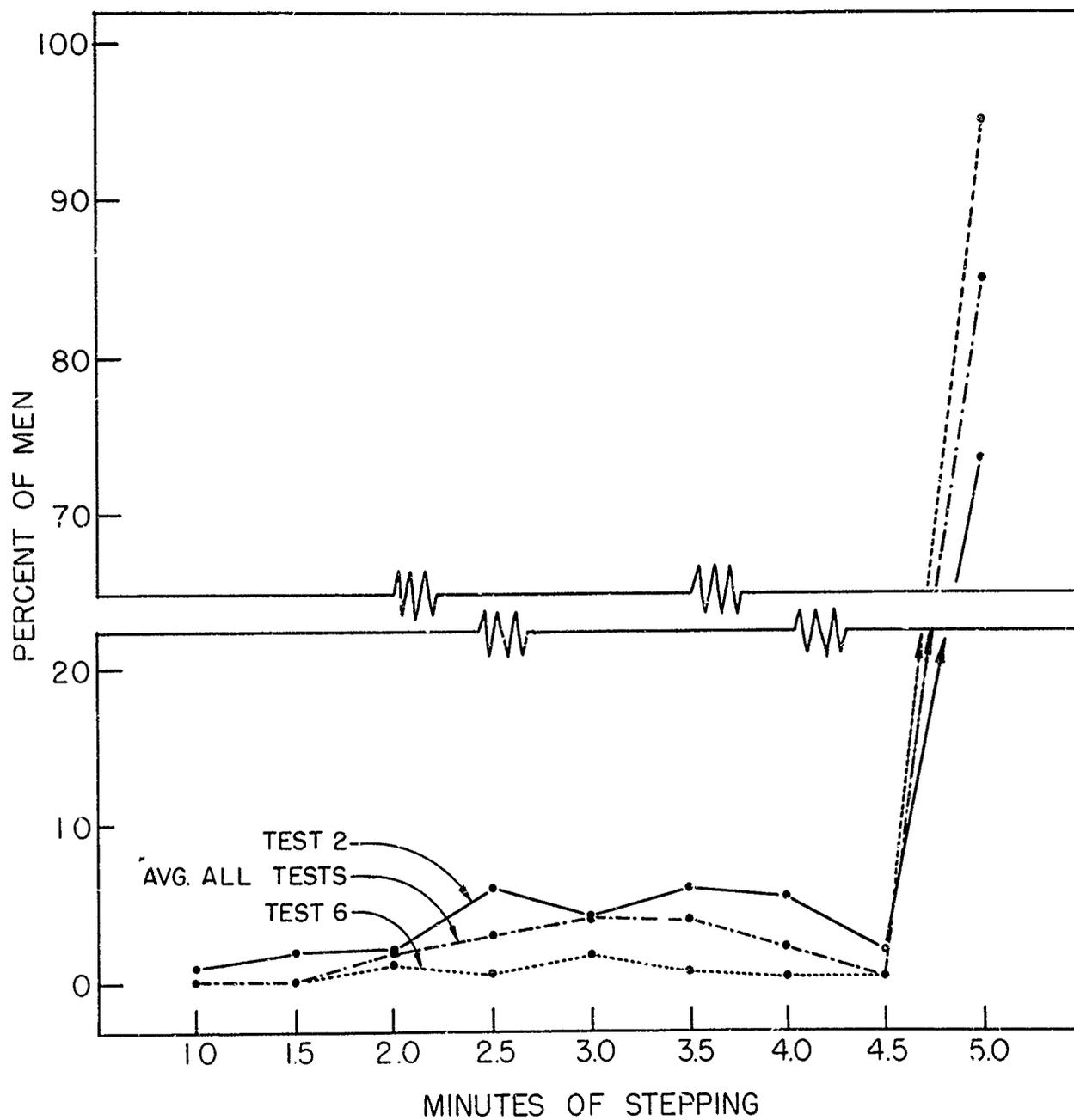


FIG. 3

DISTRIBUTION OF PULSE RATE FOR MEN COMPLETING
FULL FIVE MINUTES ON STEP TEST

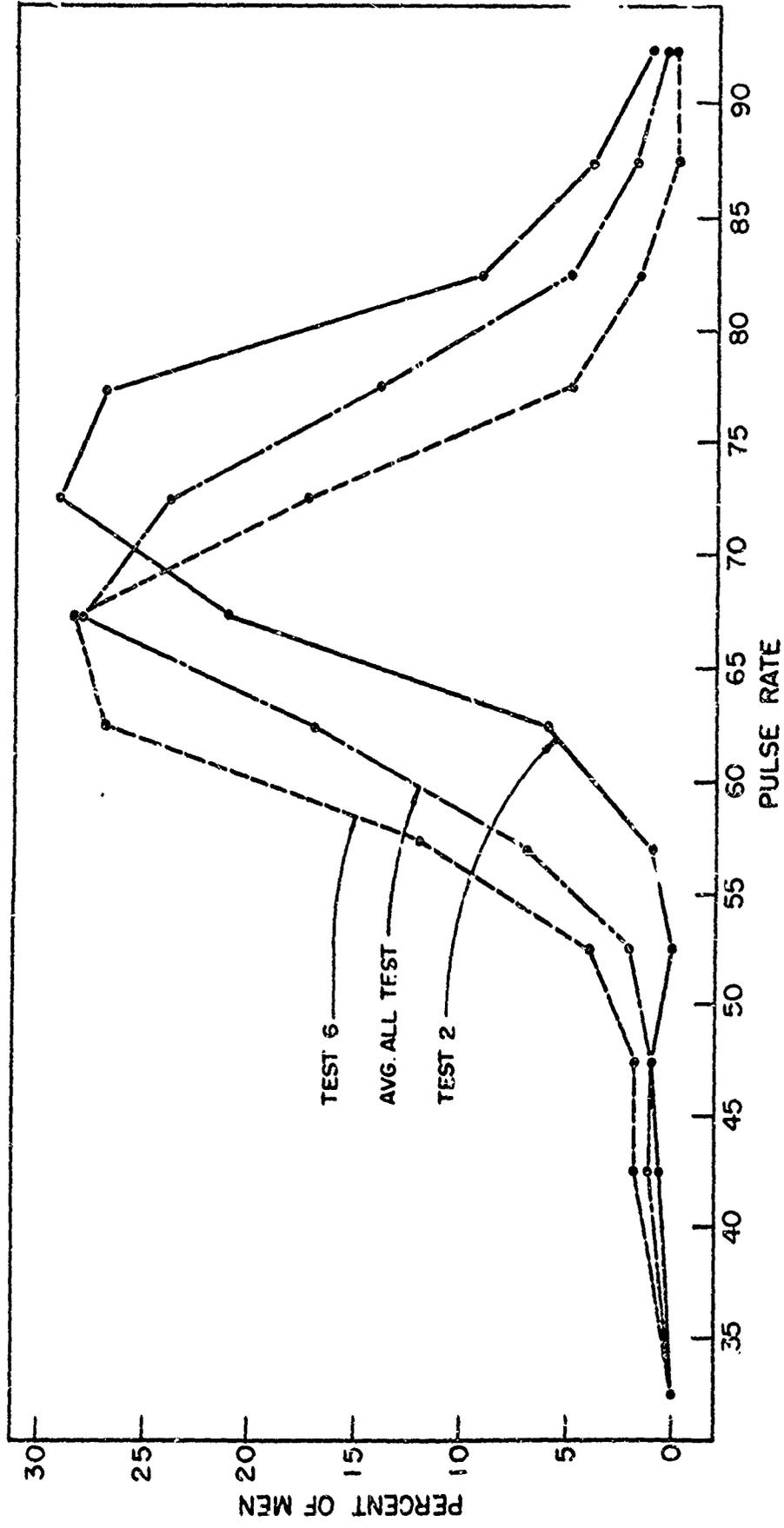


FIG 4 DISTRIBUTION OF SCORES

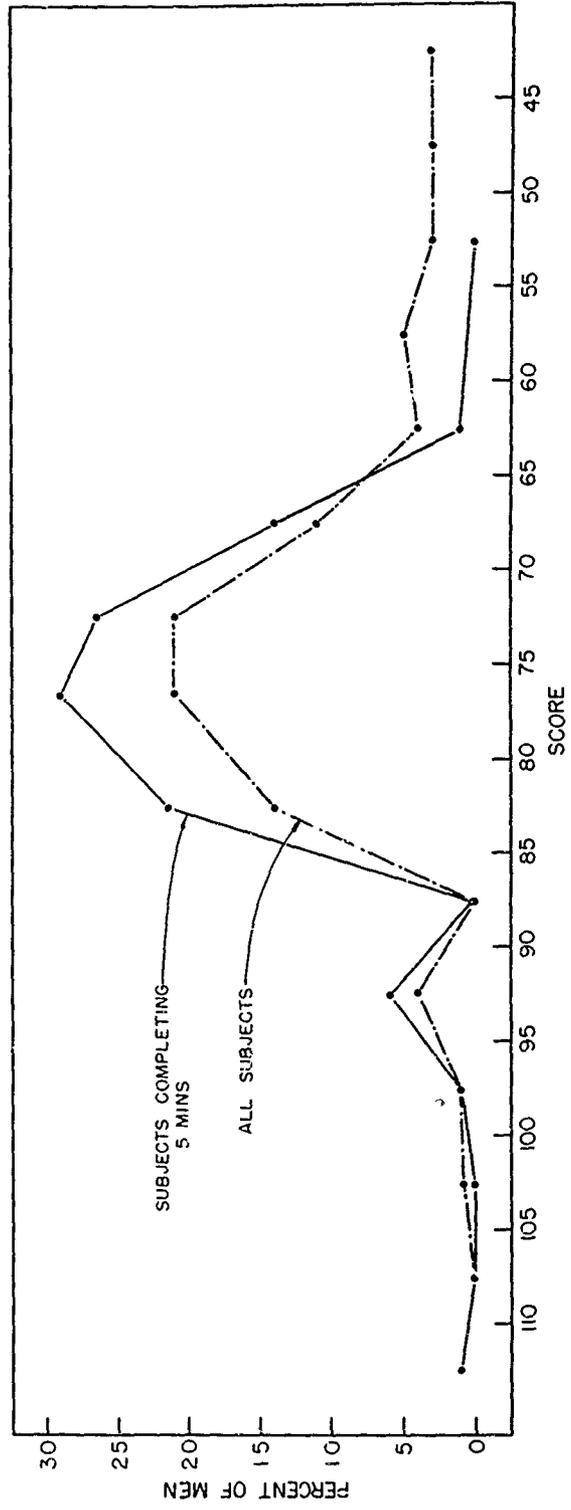


FIG 5

CORRELATION OF TIME ON AAF 300 YD. RUN WITH
PULSE RATE RESPONSE ON HARVARD TEST

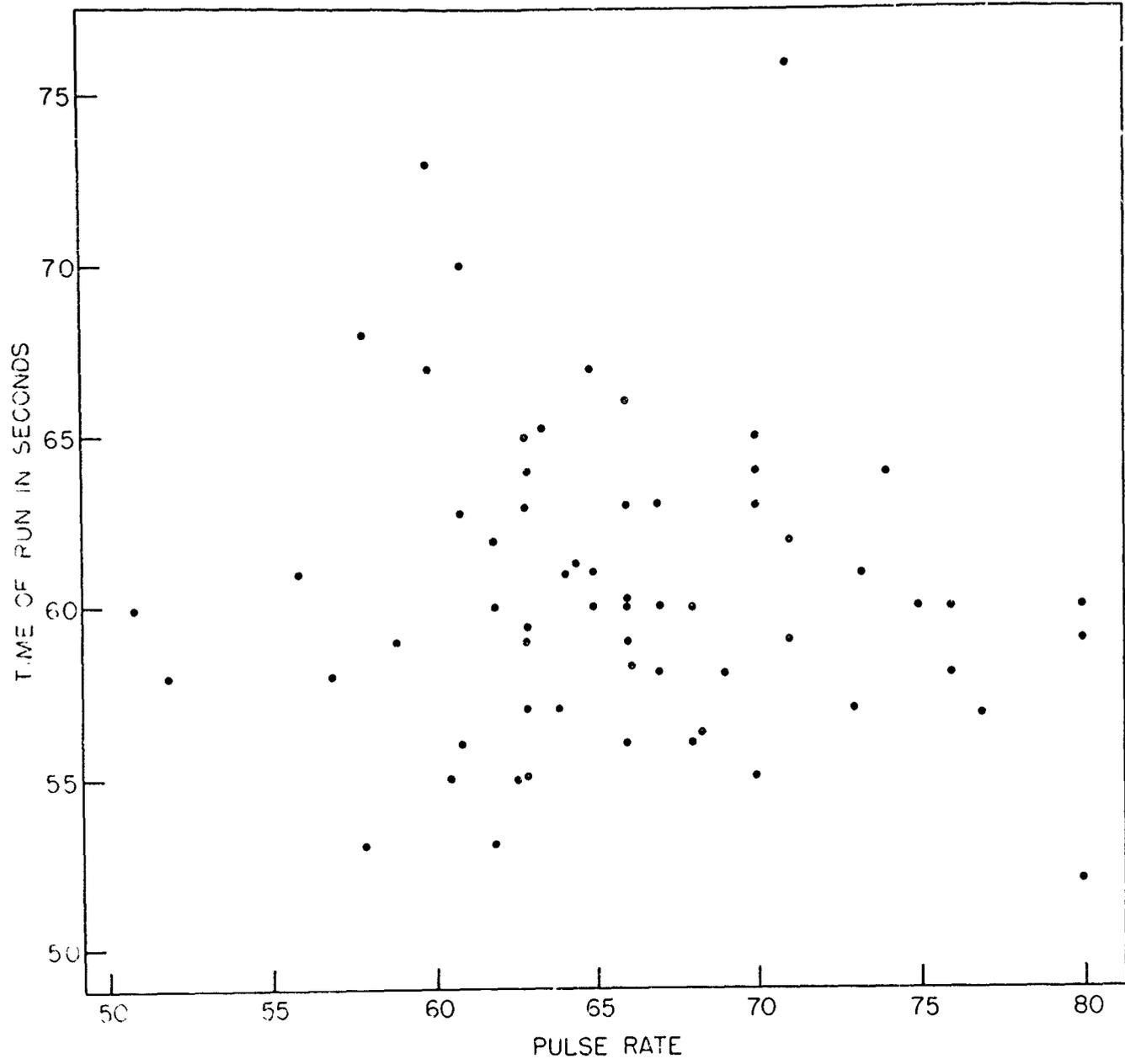


FIG 6

DISTRIBUTION OF SCORES ON STEP TEST AMONG
GROUPS OF DIFFERING FITNESS

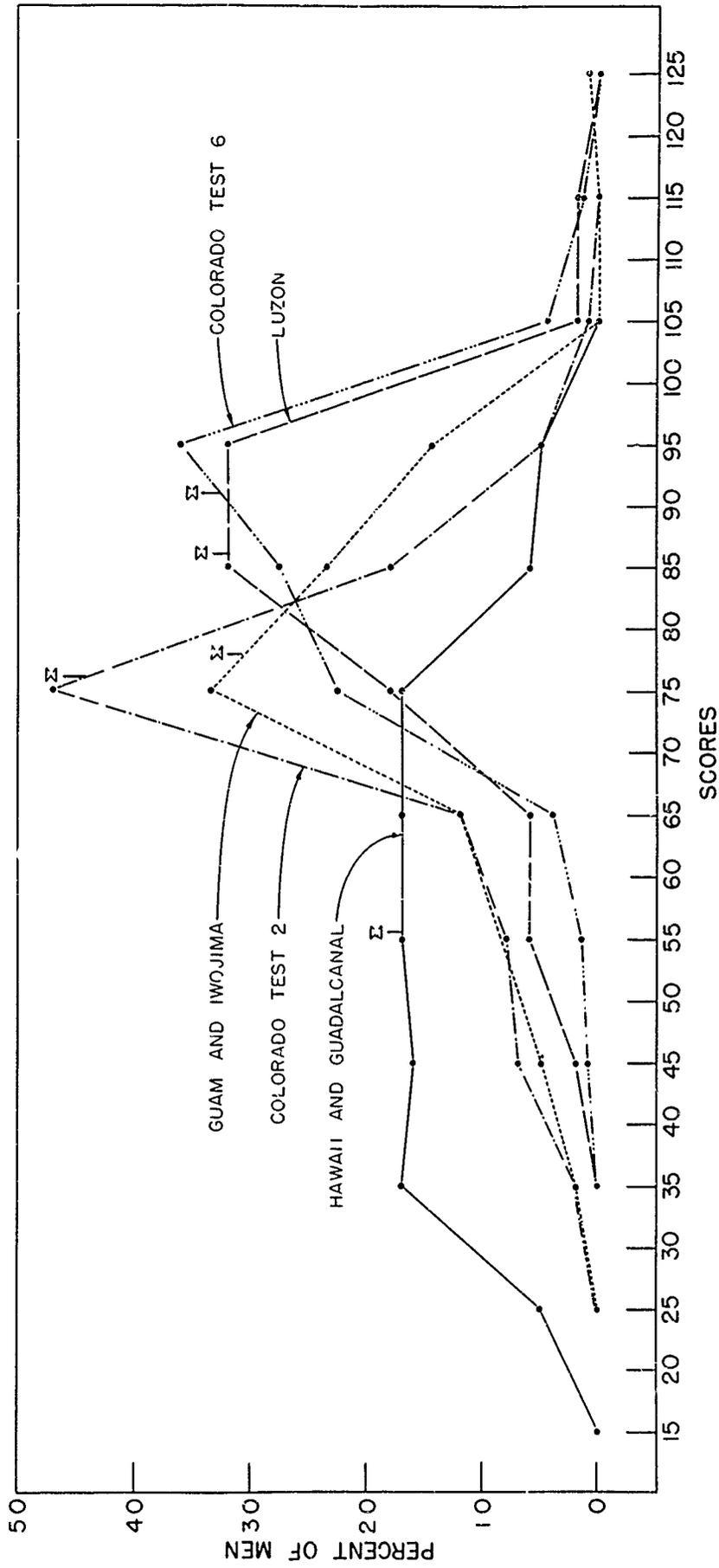


FIG. 7

DISTRIBUTION OF RUNNING TIME IN AAF
SHUTTLE RUN

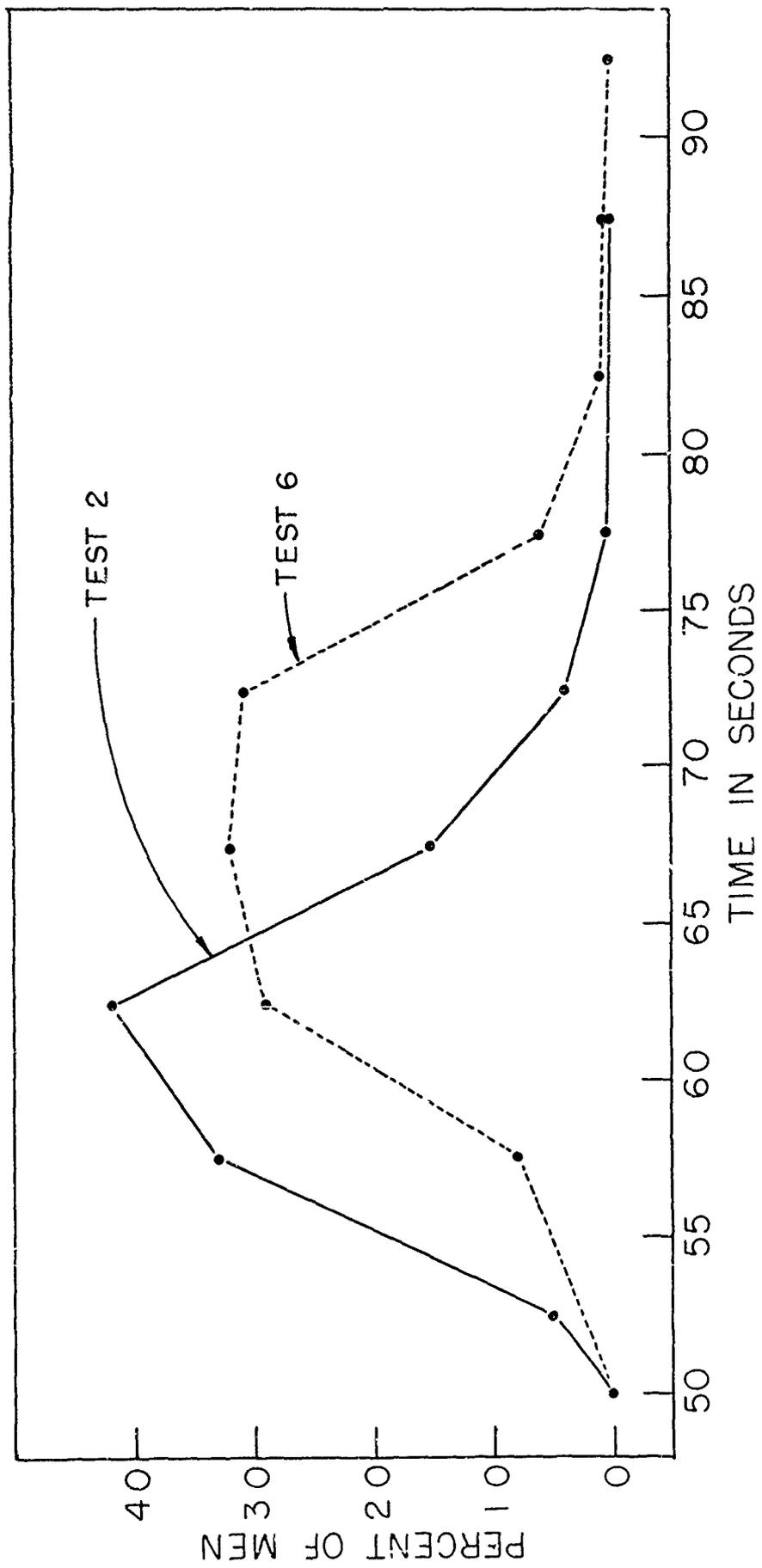


FIG. 8

SCORING CHART OF AAF SHUTTLE RUN

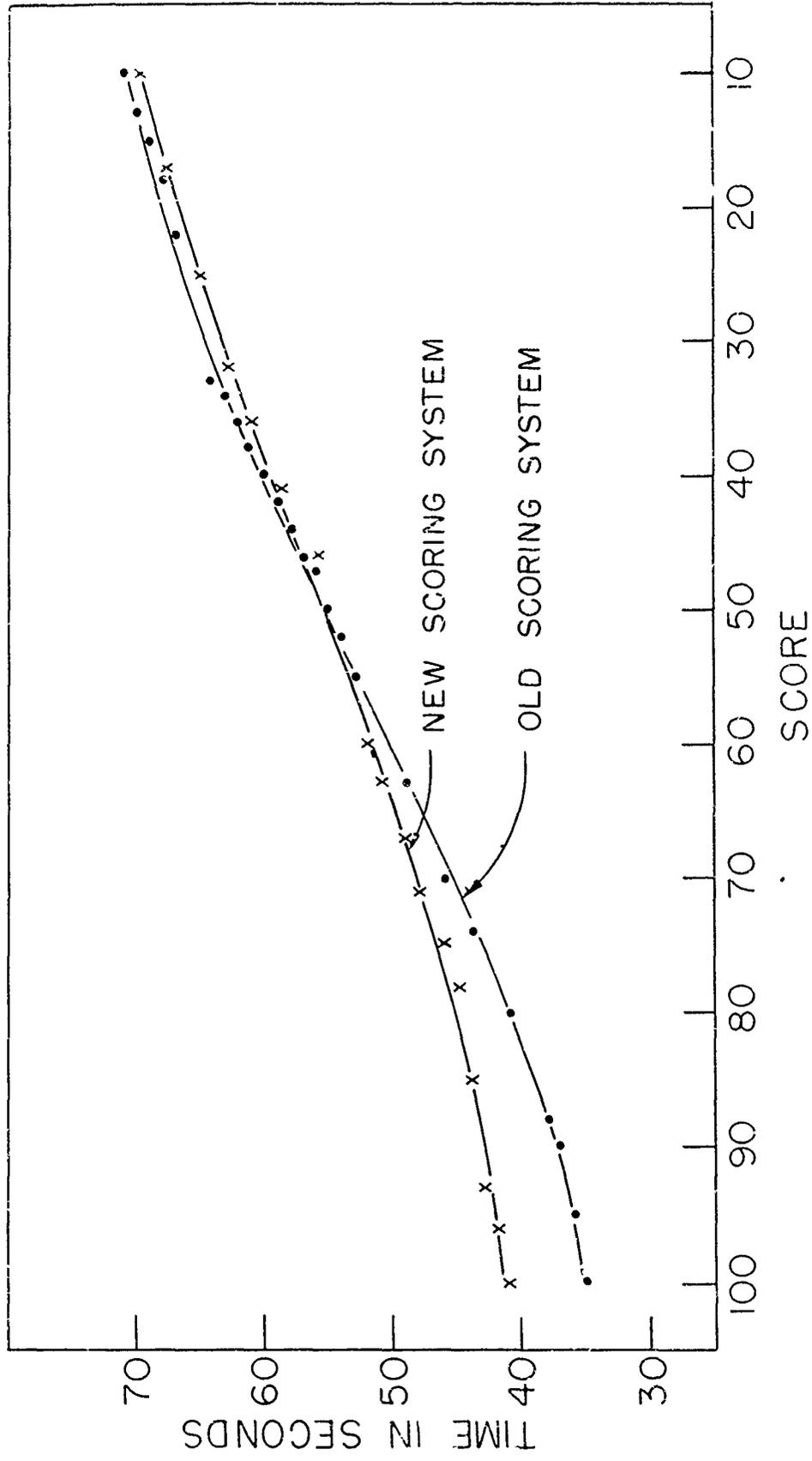


FIG. 9

DISTRIBUTION OF NUMBER OF SIT-UPS

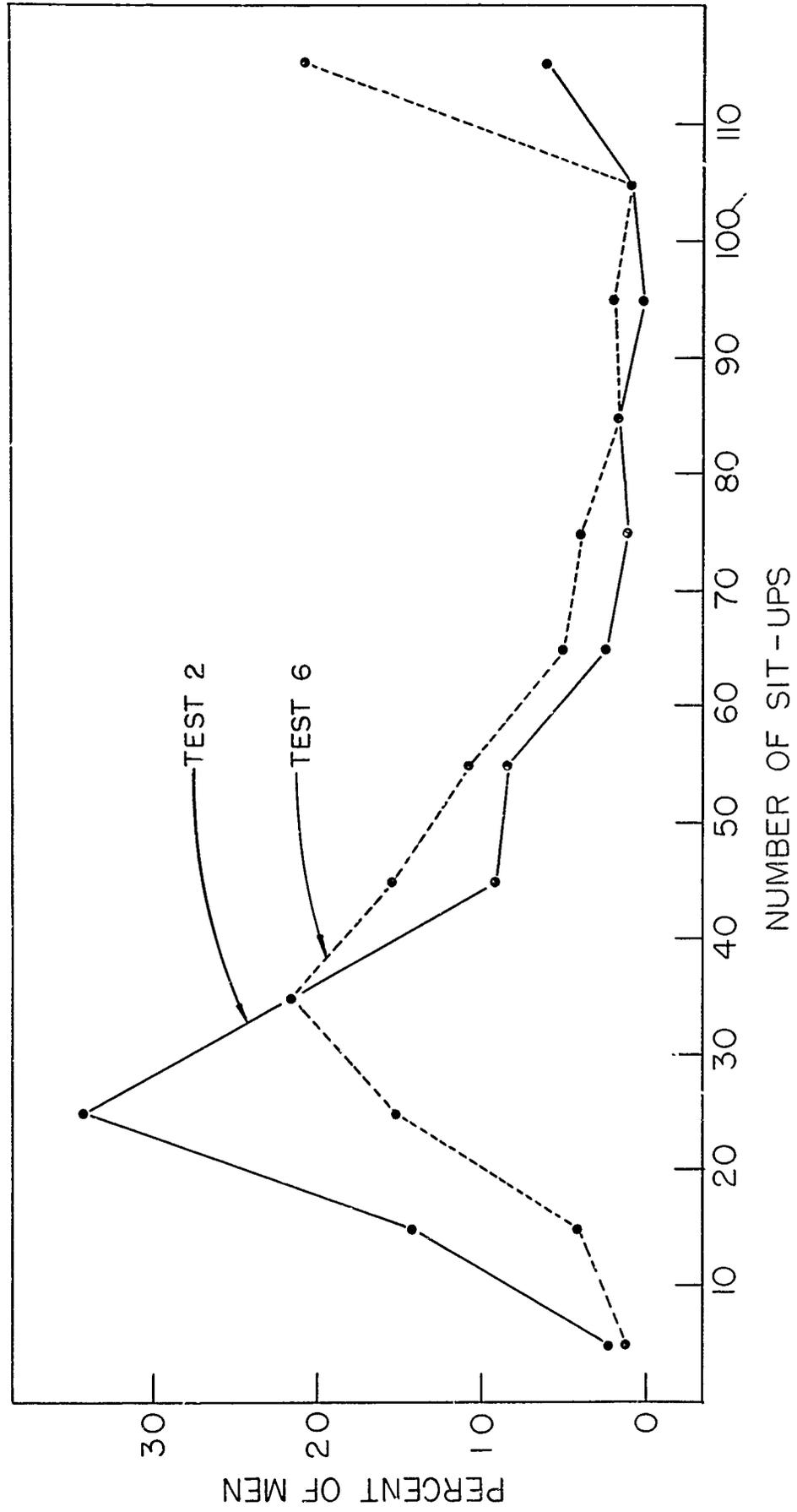


FIG. 10

SCORING CHART FOR AAF SIT-UPS

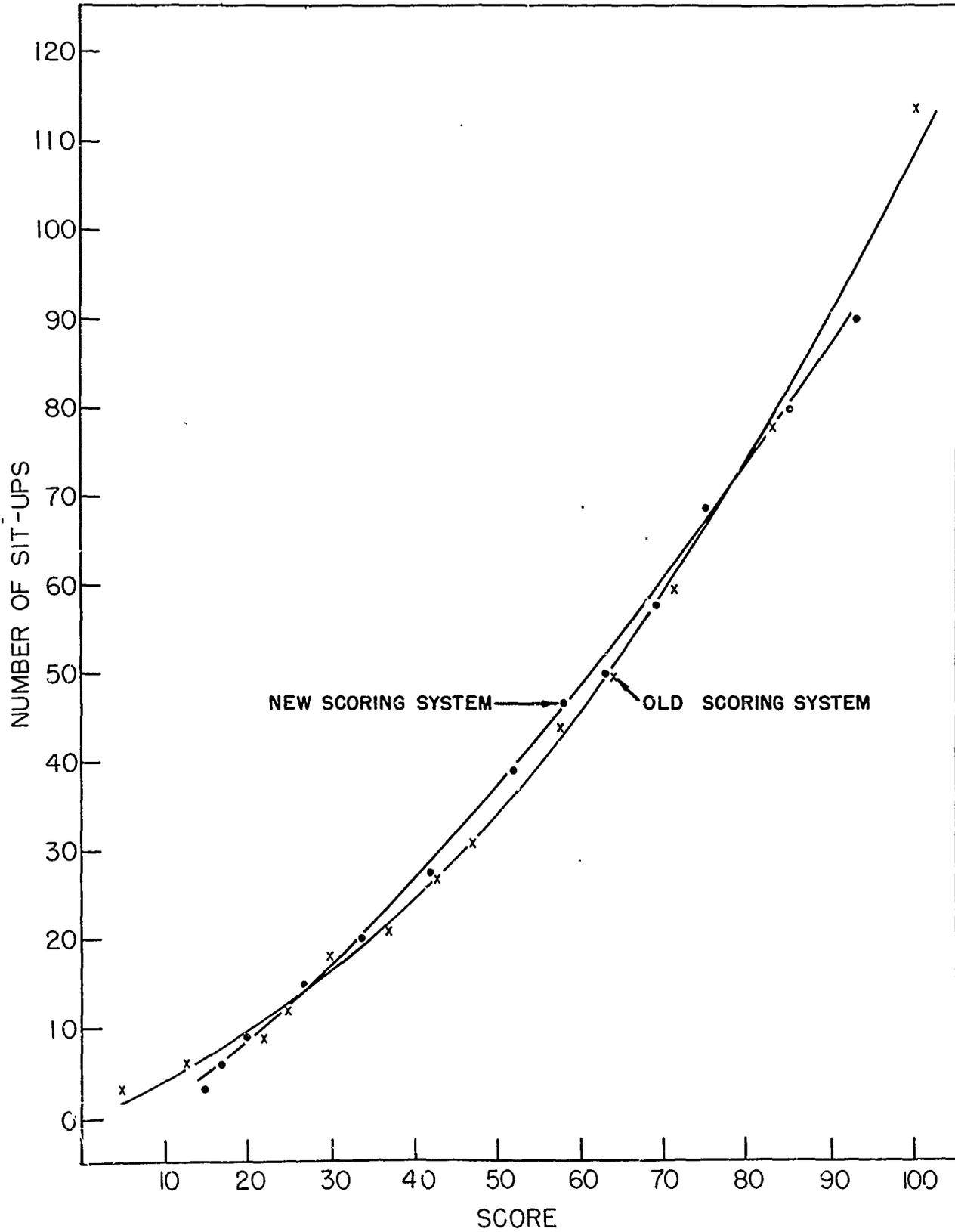


FIG. 11

DISTRIBUTION OF NUMBER OF PULL-UPS

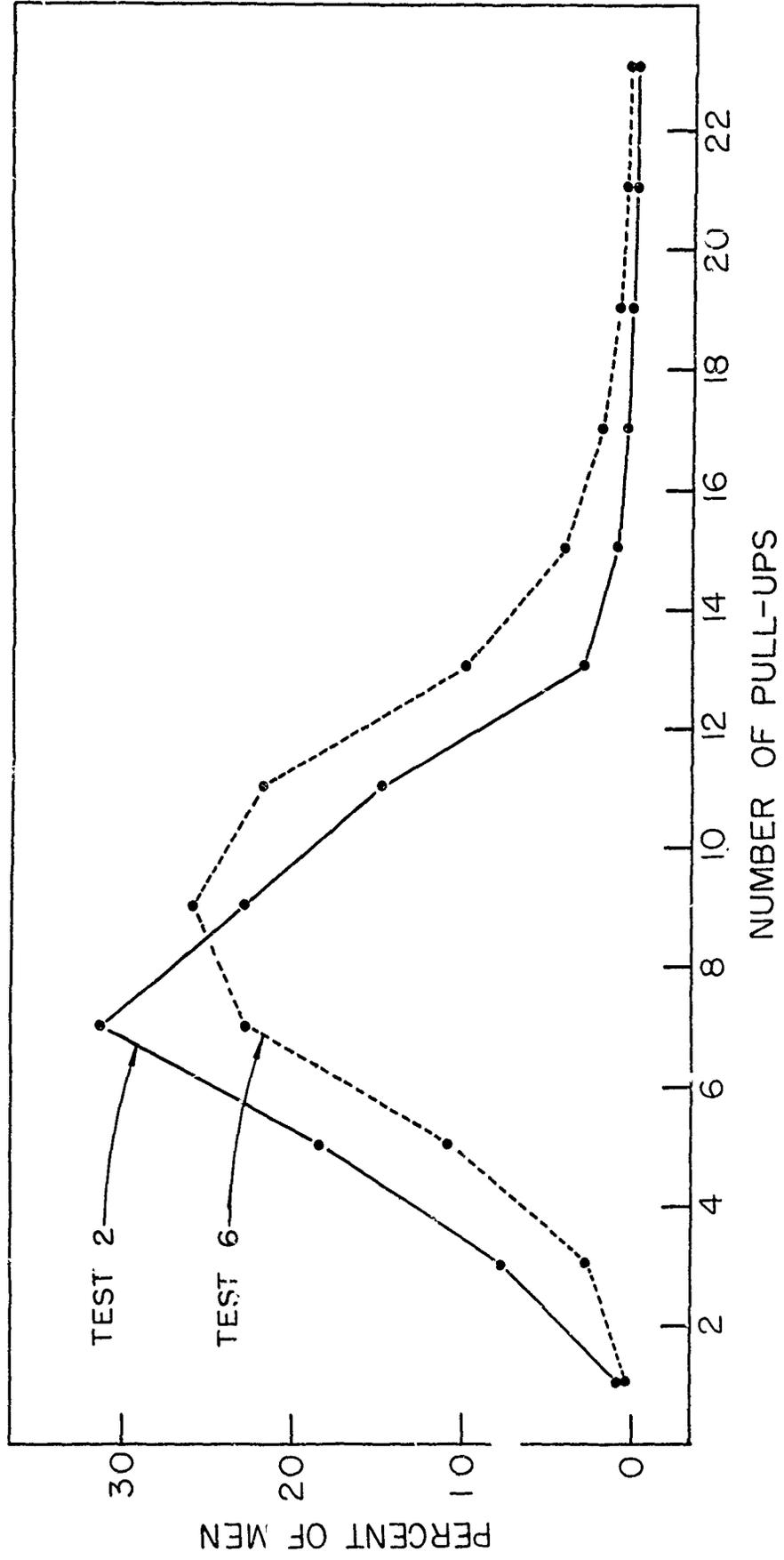


FIG. 12

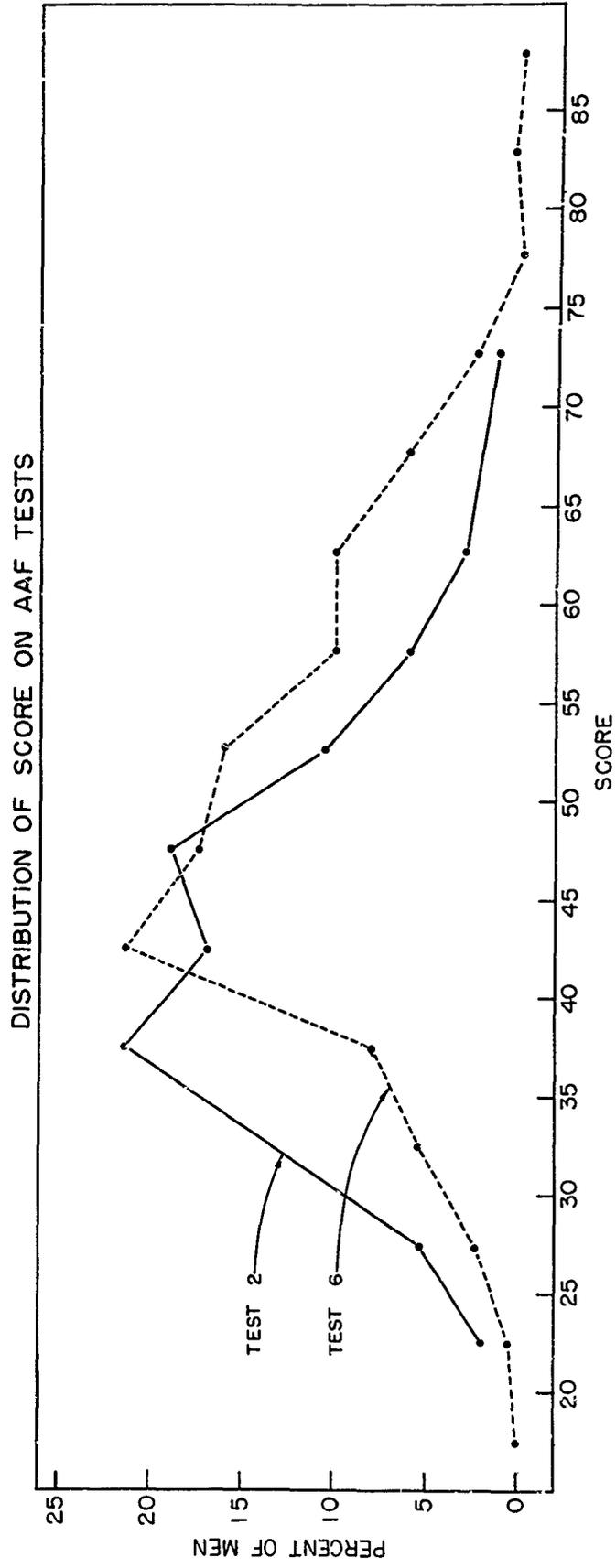


FIG. 13

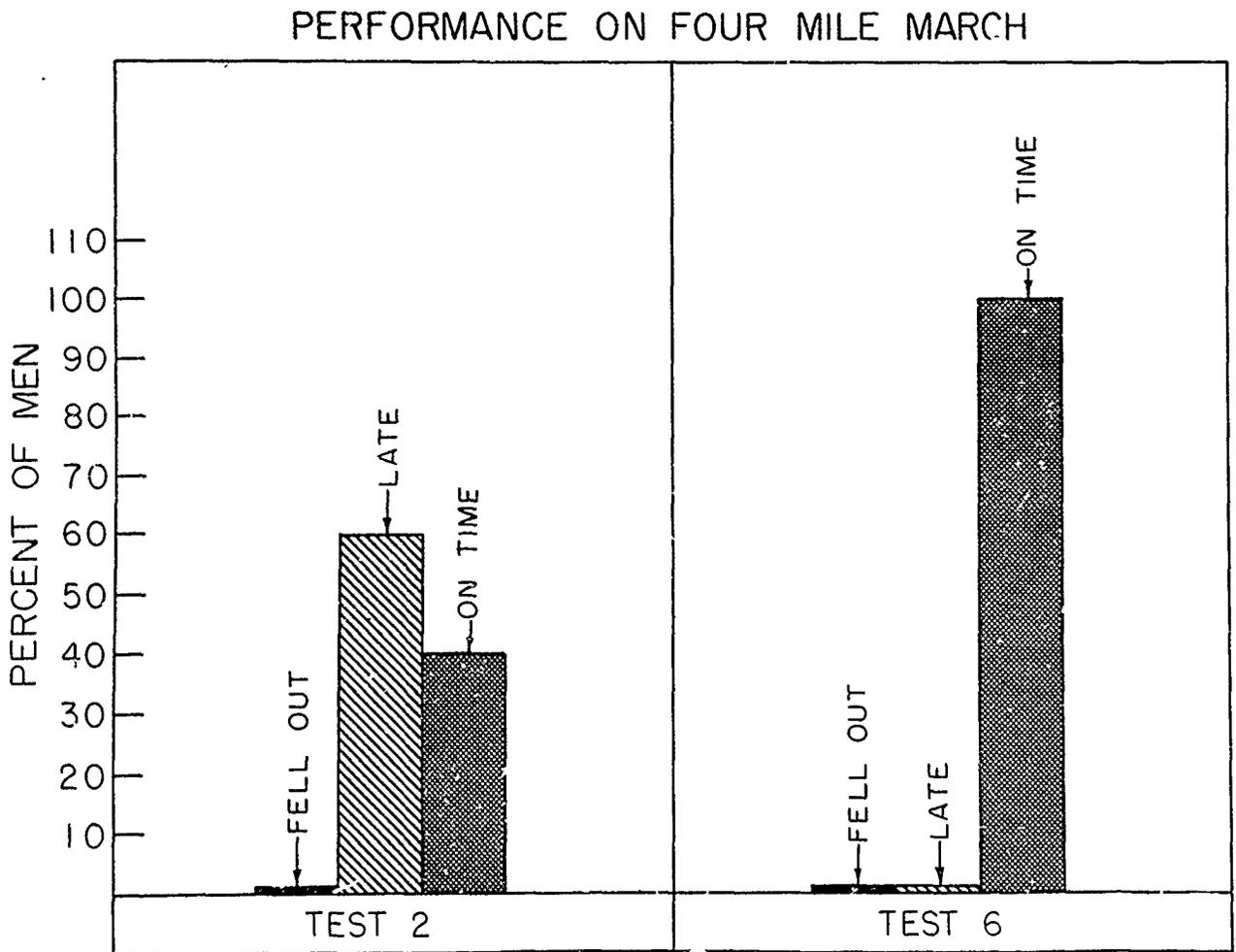


FIG 14

PERFORMANCE DISTRIBUTION ON 300 YARD RUN

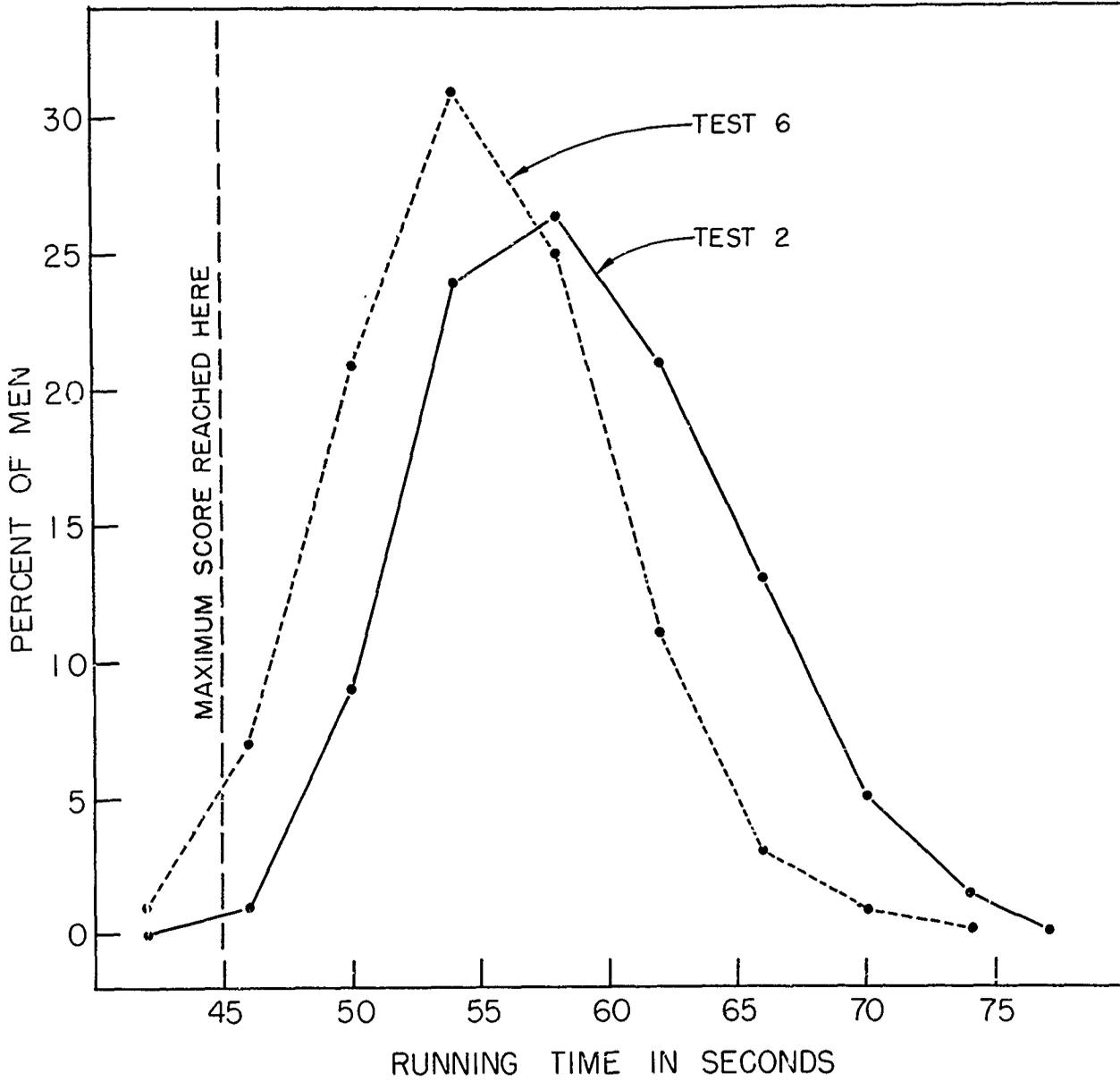


FIG 15

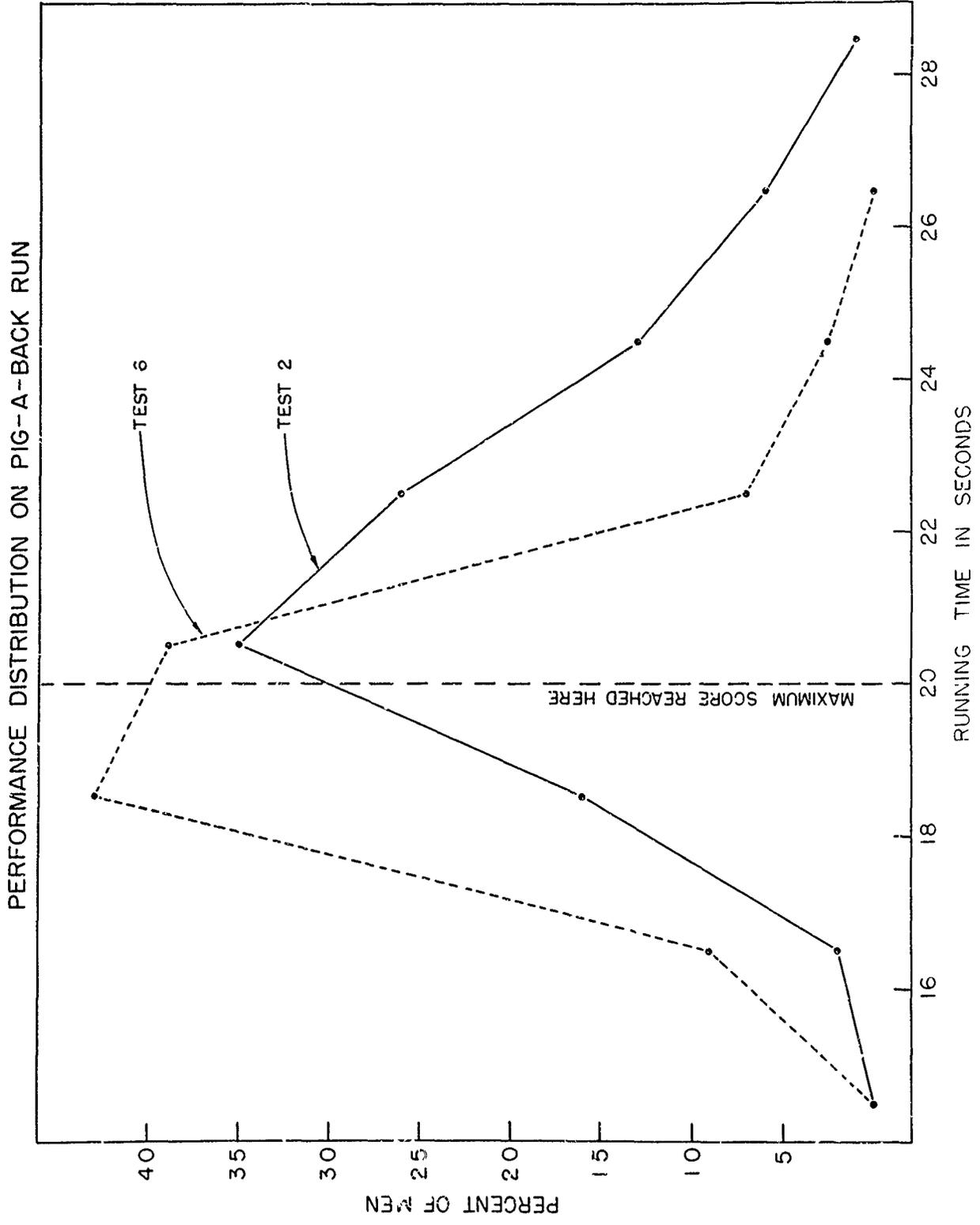


FIG. 16

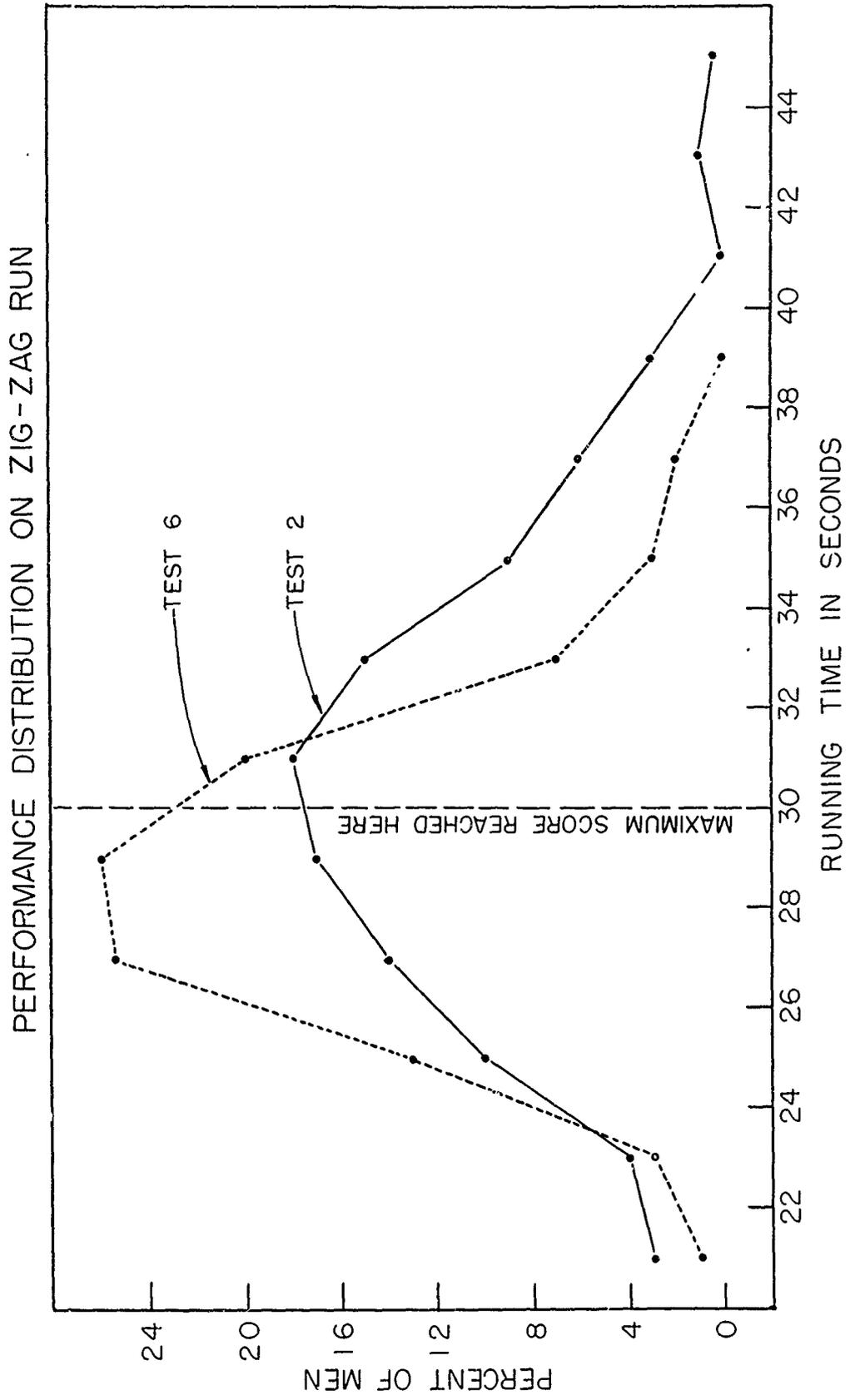


FIG.17

PERFORMANCE DISTRIBUTION OF PUSH-UPS

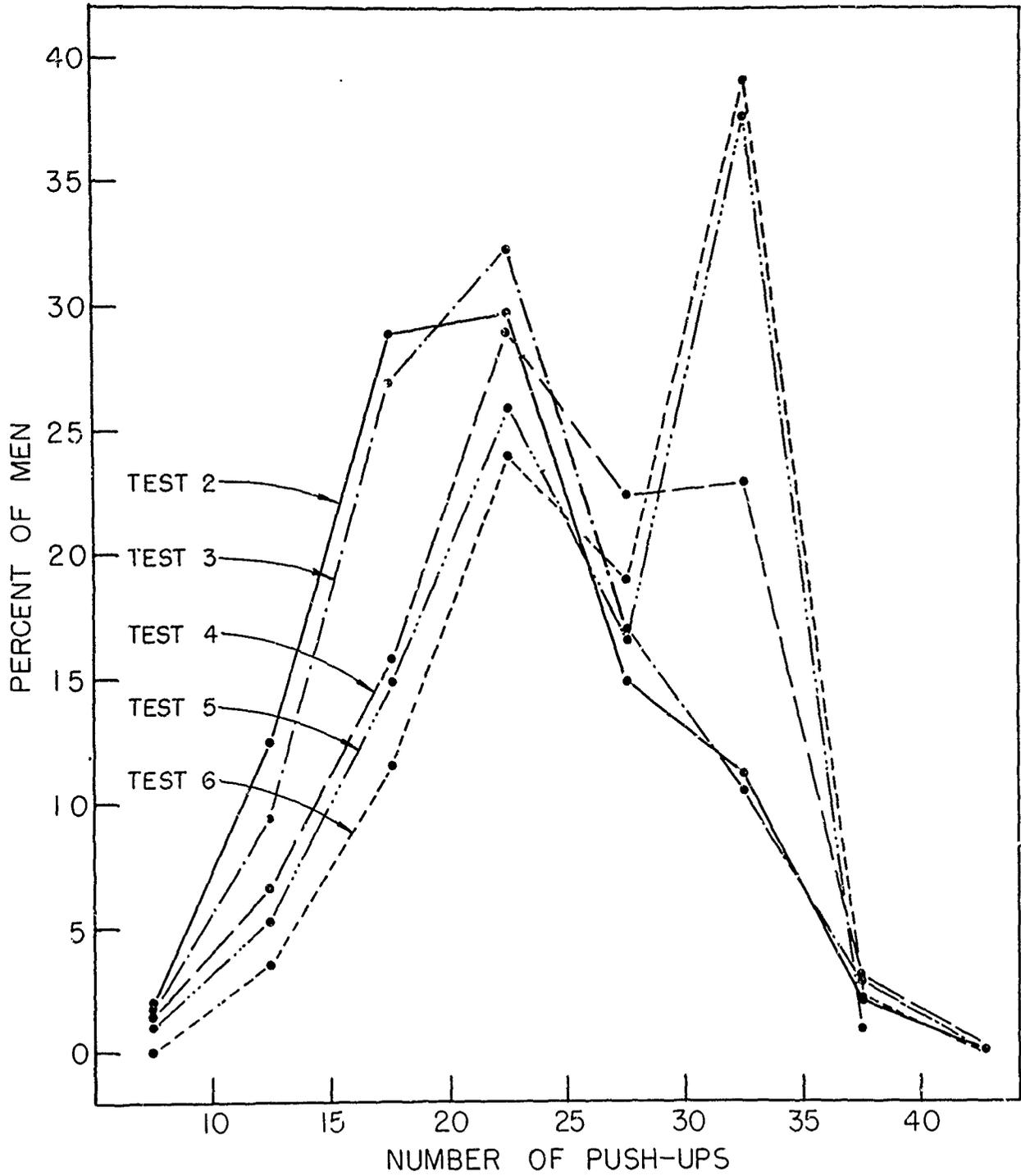
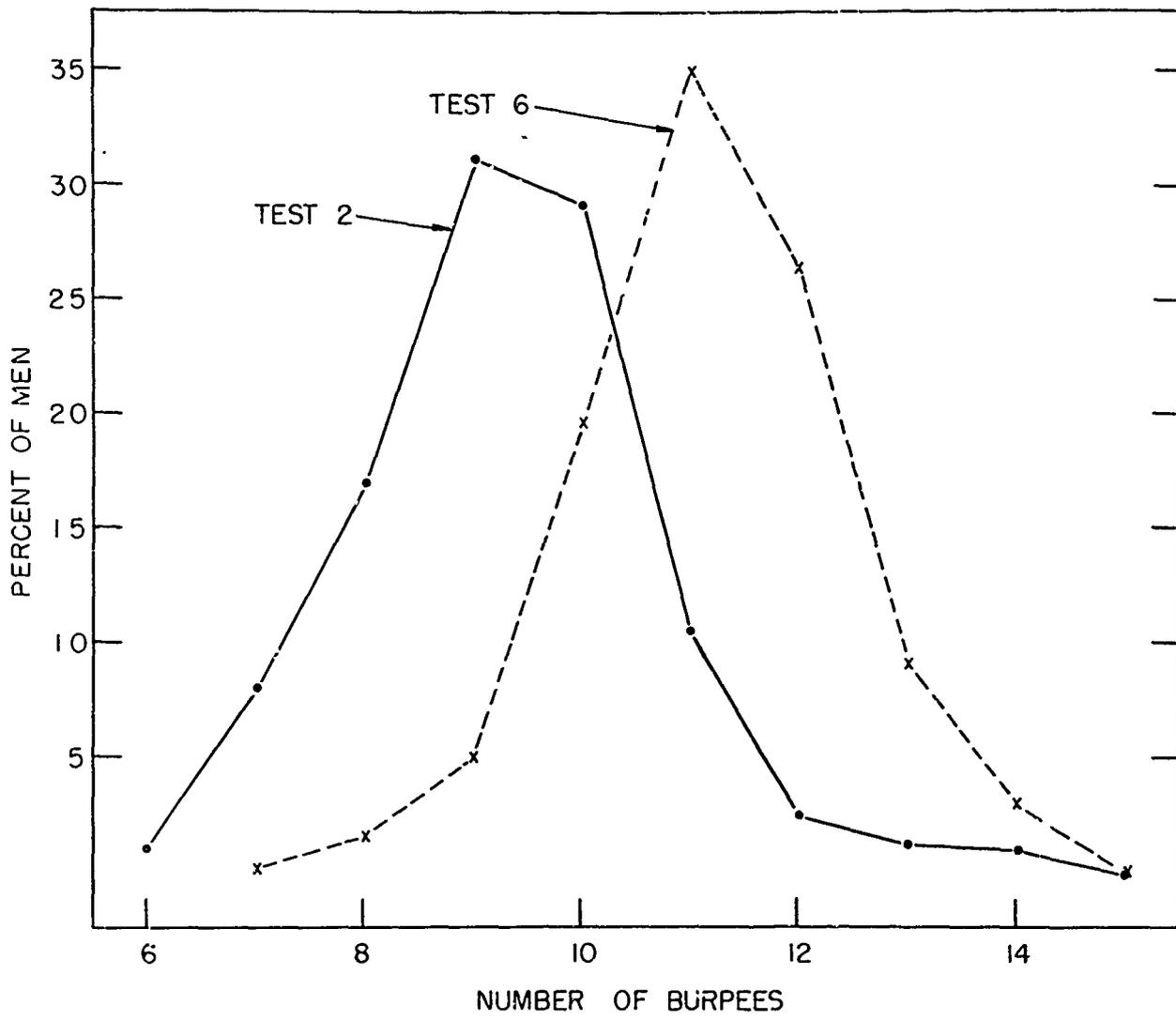


FIG. 18

PERFORMANCE DISTRIBUTION OF BURPEES



DISTRIBUTION OF SCORES ON AGF TEST

FIG. 19

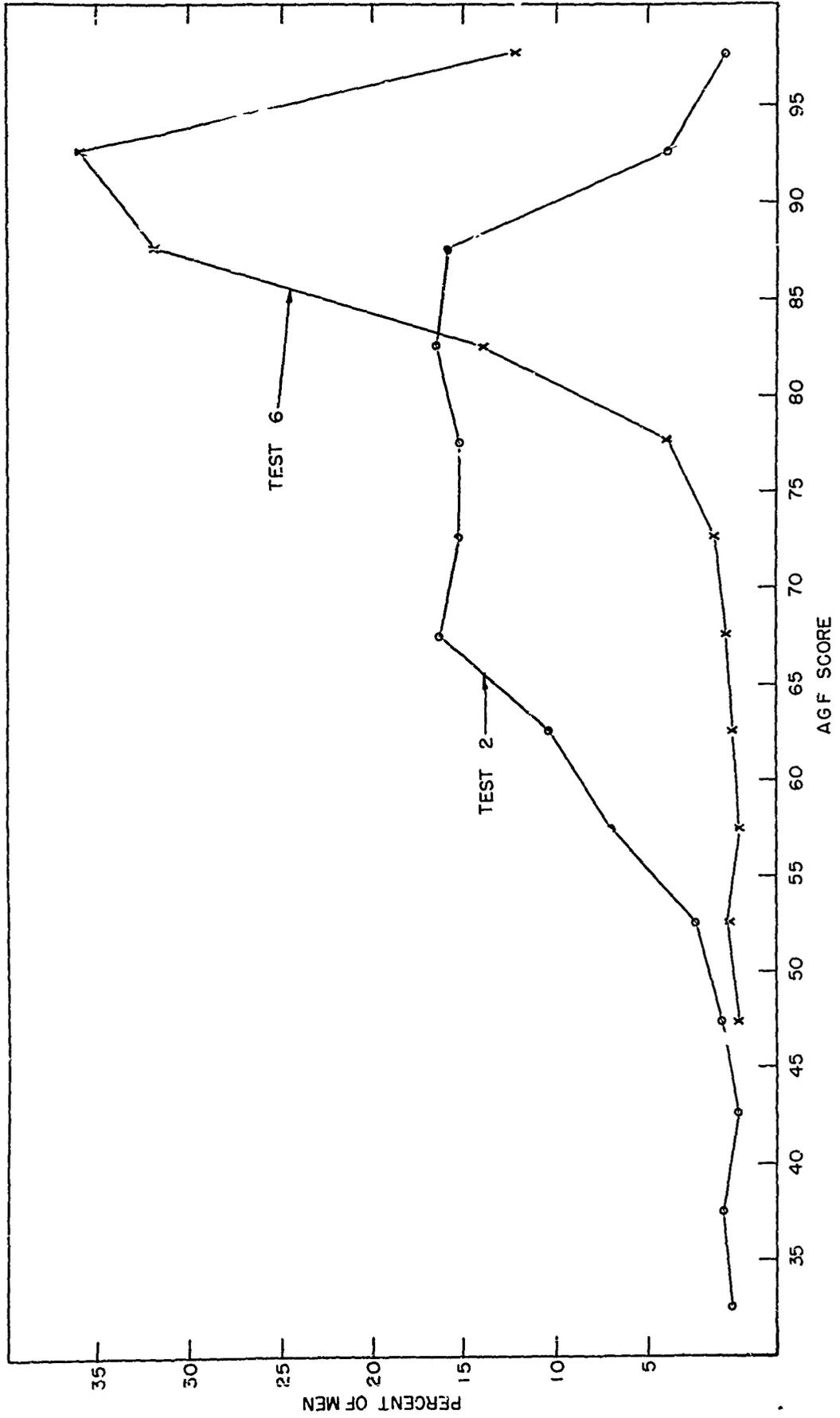


FIG. 20
THE IMPROVEMENT IN FITNESS SCORES IN EACH
TEST WITH PHYSICAL CONDITIONING

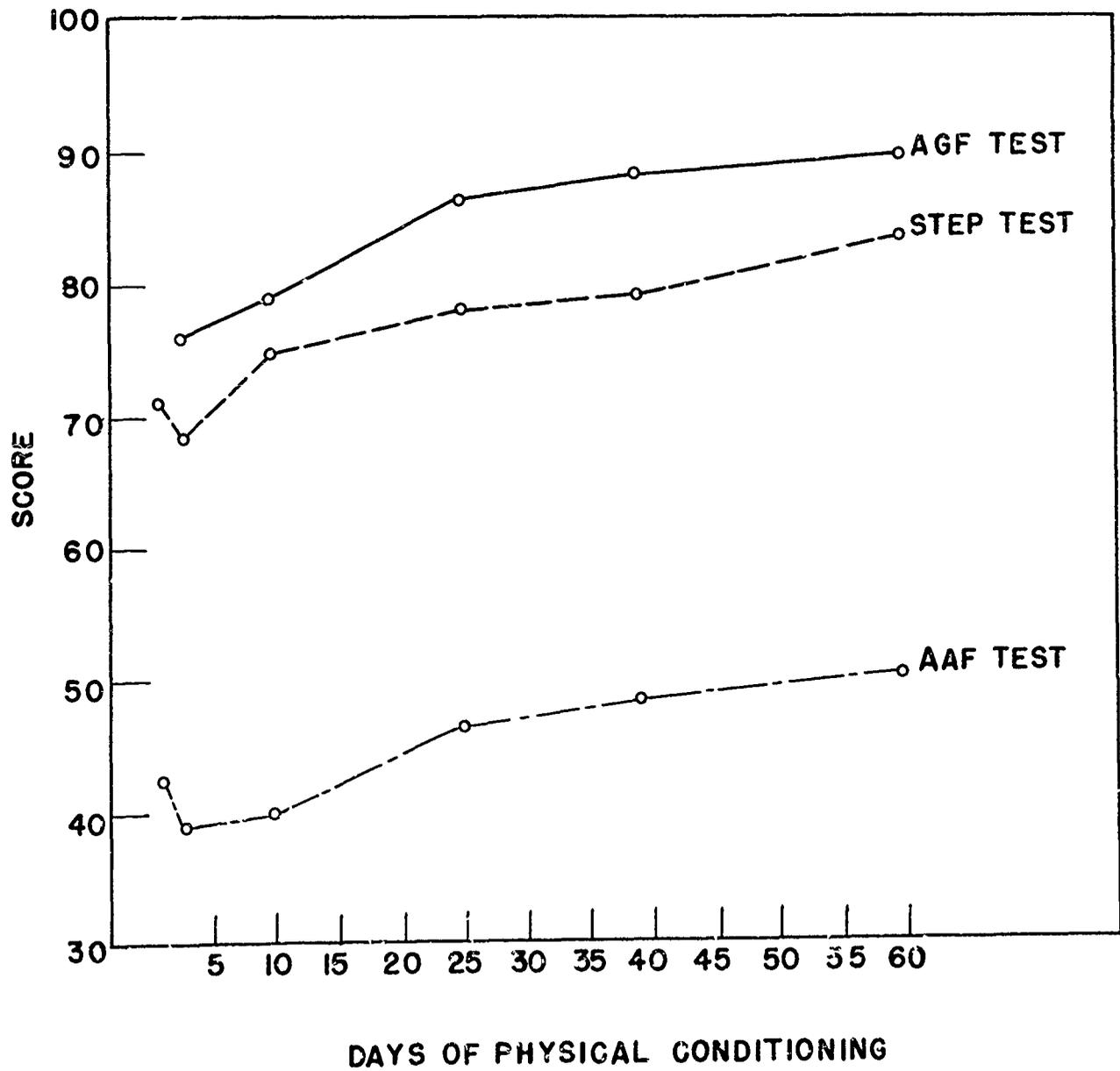


FIG. 21
 PERCENT OF MEN RATED GOOD FAIR AND POOR BY SKILLED
 OBSERVERS COMPARED TO COMBINED SCORE OF THREE
 FITNESS TESTS

