The Discovery of Aptitude and Achievement Variables

J. P. Guilford
University of Southern California, Los Angeles

THE PROGRESS MADE IN APTITUDE testing in the Army Air Forces during the recent war has already been discussed in *Science* (*2*). The brief account to follow is written with the belief that the general reader will also be interested in some of the fundamental scientific gains incidental to the AAF Psychological Program, or, more specifically, the gains in understanding the *nature* of human resources.

Under pressure of the urgent emergency during the first months of the war, the traditional approach to the development of aptitude tests for the selection and classification of pilots, bombardiers, and navigators was followed. Job analyses were made, with attempts to break the performances of men in specialized jobs into distinguishable and significant psychological functions. Tests attempting to assess those functions were constructed, and a certain degree of success was achieved. As data accumulated, however, it was recognized that the job-analysis categories were of little value in revealing why some tests were valid for the selection of good trainees, why others were not, and why some pairs of tests which appeared to be quite dissimilar exhibited substantial degrees of intercorrelation. Recourse was made to the application of factor-analysis procedures, and it soon became apparent that in this direction lay not only the answers to many puzzling questions concerning specific tests but also the general frame of reference for an enlightened test development program.

Factor theory conceives of human personality (in more operational terms, individual differences) as being economically describable by reference to a limited number of distinguishable dimensions. These fundamental variables are not readily observable by ordinary procedures because of their intricate and varied manifestations in human behavior, including that part of behavior seen under more standardized conditions, the psychological test. Tests correlate with one another to the extent that they measure in common one or more of the underlying variables or factors. From the high, low, and moderate degrees with which tests intercorrelate, the common factors can be detected by statistical operations. The attachment of psychological meaning to a factor depends upon the apparent common features of the behaviors that are symptomatic of it and upon the insight of the investigator. Several methods have been proposed for the factoring of a correlation matrix (intercorrelations among a collection of tests). Experiences of the AAF psychologists led to the adoption of the Thurstone centroid method of extracting factors, followed by the rotation of reference axes into a positive manifold and simple structure, as being the most fruitful procedure (*4*). Intelligibility and reproducibility of factors and of factor loadings in tests were thus best achieved.

It is the purpose of this brief account merely to list and to define partially the factors revealed in the AAF. Verifications of previously discovered and substantiated variables, which, in tests of aptitudes, Thurstone has called "primary abilities," (*5*) will be mentioned first. These will be followed by a list of factors which, by their repeated verification in the AAF, can be proposed as serious contenders for recognition and by a much longer list of newcomers, the existence and identities of which require further verification.

Prewar factors verified in the AAF include:

(1) *Verbal.* This is easily verified in any test that involves individual differences in the understanding of the meanings of words. Vocabulary tests are the strongest and purest measures of it—a fact not generally accepted previously. Reading-comprehension tests have their strongest loading in this factor, but they are also usually related to others in moderate or small degree.

(2) *Numerical.* The strongest and purest tests of this factor require nothing except speed and accuracy in the four numerical operations. It tends to creep into almost any test in which numbers must be used by the examinee, not only by way of fundamental operations but also in noting number size, rank order, and the like. Its exact limits are still to be determined.

(3) *Perceptual speed.* This should more strictly be denoted as "visual-perceptual speed," for it has been observed only in visual tests. The quick, yet accurate, grasping of visual details, features, similarities, and differences seems to characterize tests heavily loaded in it. Pure tests of this factor are rather easy to achieve.

(4) *Associative memory.* Thurstone named what is probably the same factor "rote memory." This is found in tests requiring the memorizing of elements—letters, words, numbers, pictorial objects—in pairs and a later associative recall. The elements may be meaningful or nonsensical.

(5) *General reasoning.* This term is an admission of failure to identify more exactly the factor which appeared in a larger number of reasoning tests than any other reasoning factor. The test that most consistently leads in this factor is that of arithmetic reasoning. It is not a mathematical-reasoning factor, however, for it is found

in nonmathematical tests. In the AAF results there is little to support the traditional distinction between inductive and deductive reasoning abilities. No reasoning test was pure in any factor. Since all tended to be factorially complex, definitions of the three reasoning factors could not be achieved. One hypothesis is that the general-reasoning factor represents a diagnostic ability—the ability to grasp the nature of problems.

The next factors are fairly well verified. Only one (visual memory) has a clear prewar counterpart.

(6) *Spatial relations.* One of the most significant AAF findings was the separation of this factor and the next one, called *visualization.* It seems quite clear that prewar results had confused the two. Thurstone called his primary ability in this area "space," but he defined it as if it were visualization (*5*). The space factor, as here defined, is found in psychomotor tests as well as in printed tests. It seems to be a perceptual awareness of the arrangements of objects with respect to right-left, up-down, and out-in dimensions. Correct choice as to direction of movement in some psychomotor tests depends upon it.

(7) *Visualization.* This factor, which should probably be modified to "manipulatory visualization," is prominent in tests requiring one to imagine transformations, movements, or other changes. A typical test presents a picture of a square paper being folded one or more times, in successive steps, and a hole of a given shape being cut out. The examinee selects one of five other pictures, each showing plausible creases and holes, as the paper might appear after being unfolded.

(8) *Mechanical experience.* The AAF analyzed almost every type of so-called mechanical tests. Printed tests often designated as "mechanical" showed, depending upon the test, moderate loadings in perceptual speed, spatial relations, and visualization. Many of them showed loadings in a factor which compelled the conclusion that its existence depends heavily upon learning —hence the factor name, "mechanical experience." Tests of mechanical information (stressing knowledge of tools and automobile parts and their functions) are almost unique in this factor. The popular term "mechanical aptitude" covers a rather loose collection of things and, when used, should be carefully scrutinized. Mechanical jobs vary all the way from tinkering with clocks and operating a steam shovel to designing a bridge. Each job, though called "mechanical," probably has its unique pattern of requirements in terms of human resources.

(9) *Length estimation.* Probably the best test of this factor is one requiring a simple and direct comparison of the lengths of lines. The factor also appears in tests involving objects of more than one dimension. Its degree of generality has not been determined.

(10) *Visual memory.* This factor, which seems to be a purely reproductive type of memory for visual patterns, can probably be identified with a similar factor found by Carlson before the war (*1*). It is found in recognition tests as well as recall tests and might be regarded as a second and independent type of visualization. Both this and the manipulatory-visualization factor would be expected to depend upon the visual cortex, and in view of their independence, one should look for two distinct corresponding properties of that cortex. In this connection, one is reminded of the distinction made in two types of eidetic imagery—rigid and plastic.

(11) *Judgment.* Job analysis, of the pilot's work in particular, stressed the paramount importance of what is popularly known as "common-sense judgment." Tests called "practical judgment" were therefore constructed. These presented items in the form of verbally described predicaments such as those a service man might encounter, each with five more or less plausible solutions, one of which was regarded as the wisest under the circumstances. Analysis of such tests showed significant loadings in the verbal factor, in general reasoning, and in the mechanical-experience factor. (Knowledge of common tools and materiel was apparently useful to the examinee in this kind of test.) Over and above these variables common to the practical-judgment tests, there was additional communality which could be called judgment. While it was characteristic of judgment tests as a class, it was also common to tests calling for practical estimations of sizes, times, and distances in everyday situations and to some tests of planning. It may actually be a fourth kind of reasoning—that is, reasoning of a judicial or critical type. It is probably a factor that is almost entirely missed in most intelligence tests, which supports many a layman's view that common-sense judgment is not covered by the IQ.

(12) *Psychomotor coordination.* This factor is common to various psychomotor tests, including those requiring coordinations of fingers and hands (finger dexterity), of arm and shoulder (rotary pursuit), and of arms and legs (using mock airplane controls). It is doubtful whether it can be identified with the agility factor found in such physical education tests as the dodge run. It is quite possible that there is a coordination factor other than this one which is restricted to the finer finger movements.

The remaining factors are merely suggestive of worth-while hypotheses.

(13) *Psychomotor precision.* This was found to be common to a finger dexterity test, a discrimination-reaction time test that required rather accurate aiming at the reaction keys, and the pass-fail criterion in bombardier training. It is hoped that the naming is more than a figure of speech in its association with precision bombing!

(14) *Psychomotor speed.* This appeared in only two rather unique paper-and-pencil tests in which the rate of marking an answer sheet was important. It strongly

suggests a prewar factor in similar tests, such as the speed of making "gates" (tally marks). Little more can be said of it at this time. Hypothetically, it would be important in the more simple, repetitive tasks of motor performance.

(15–16) *Reasoning II and reasoning III.* These two factors arose rather weakly in a number of factorially complex tests. Clues for hypotheses are seriously lacking. Reasoning II is prominent in analogies tests and the hypothesis of "reasoning-by-analogy" is tempting, but it also appears in other tests not obviously having this quality. Reasoning III is strongest in two tests, one of which seems to call for inductions (seeing systems in arrangements of lines) and one for deductions (deciphering a code). The finding of at least three distinct reasoning factors suggests that the term "reasoning" conceals a number of unrecognized functions. Some of these may be primarily biologically determined, but others may be in the form of reasoning habits brought about under educational pressures, formal or otherwise. Until the variables are better identified in this area, studies of training in thinking and of the problems of formal discipline cannot be adequately effected.

(17–18) *Space II and space III.* These factor names are not only very general but also highly tentative. Space II is prominent in two of Thurstone's space tests involving the ability to recognize from pictures of human hands held in varied positions whether each picture is of a left or a right hand, in the one, and the ability to tell whether two U. S. flags, as pictured, show the same or the opposite sides, in the other. One hypothesis is that kinesthetic imagery plays a role in the solution of such items for some individuals. There is no suitable hypothesis to offer for space III.

(19–20) *Memory III and memory IV.* A third memory factor was strongly common to two tests. The task was one of paired associates in which names were memorized for objects: airplanes in one test and outline lakes, rivers, and bays in the other. These tests also had in common the more general associative-memory factor previously mentioned. This fact, taken together with that of several reasoning factors, suggests the possibility of hierarchies of factors within an area of mental functioning, with different levels or different degrees of generality, or both. It would seem that, although such functions are statistically independent or at least separable, they quite commonly operate conjointly in behavior. Memory IV is named as such with considerable hesitation. It could be defined as memory for verbal instructions or, generalizing somewhat, as memory for verbal content. This factor was found in a small number of tests having in common intricacy of instructions and tasks. Forgetting any part of the directions while attempting an item might well result in failure on the item. The tests were designed to see how well the prospective pilot could "keep in mind" a number of things that had to be taken into consideration in performing a task. One memory test (called Memory for Tactical Plans), which consisted of a quizzing two hours later on a briefing given orally for a mock military mission, although not analyzed along with the tests just mentioned, was known to have an unknown factor or factors related to the criterion of pass-fail in pilot training over and above factors already accounted for. It is a reasonable hypothesis that "memory for instructions" is common to this test and those known to be loaded in memory IV.

(21–22) *Mental set I and mental set II.* The first is strongest in tests of following directions when these keep changing as the examinee proceeds. There is little for the examinee to remember for very long. The significant variable seems to be an adaptability to changing rules. It is not a matter of flexibility of set, in a sense opposite to perseveration. Explicit attempts to bring out an expected common factor of perseveration in a number of tests definitely failed. Mental set II seems to represent a breadth-of-set quality. Attention to details, overlooking none, seems relatively important. The tests in this cluster are tedious, exacting tasks. Whether definition of the factor should emphasize breadth or care in handling details is an open question.

(23) *Carefulness.* This and the next factor seem to be temperamental or motivational variables rather than abilities. It was hypothesized that navigators must exercise extreme care with details in their work, for small mistakes become magnified in terms of miles-off-course and errors in estimated time of arrival. Four tests of a complex, clerical type were devised to study this hypothesis. Analysis of the scores based upon the number of items correct yielded some of the already familiar factors in significant amounts—number, visualization, and space. The scores based upon the number of errors made, however, had in common a new factor which could very reasonably be called carefulness. Its possible relations to mental set II have not been investigated. This might be denoted as "meticulousness," a personality trait that has already been given some attention by clinical psychologists. Of great general importance in this finding is the fact that error scores and correct responses may measure something very different. It has generally been assumed that "rights" and "wrongs" in a test measure the same functions in reverse direction. The finding opens the door to a promising line of new research.

(24) *Pilot interest.* From the finding of this unity in behavior, one is led to the conclusion that our culture has molded an interest variable centering around piloting an airplane. It seems not to be a general aviation interest, and no similarly crystallized interest variables for navigation or for bombardiering were noted. The pilot-interest factor was revealed in general-information tests (which included items on knowledge

of airplanes and their operation) and in a biographical-data test score that was valid for pilot selection. This test called for facts concerning the examinee's previous flying experience, hobbies, and the like.

(25) *Mathematical background.* This is common to the biographical-data test as scored for the selection of navigators and to mathematics-achievement tests (including algebra, with some trigonometry and analytical geometry). It has no relation to numerical-operation tests or to arithmetical-reasoning tests. Like the mechanical-experience factor, it seems to be a culturally determined unity in individual differences.

(26) *Social-science background.* This is mentioned with extreme hesitation because it was revealed in conjunction with only two tests—information tests in history and geography. The only other supporting evidence, (and it is quite slender) is that this factor correlates slightly negatively with the pilot criterion, which is also true of biographical indications of specialization in the social sciences. This hypothesis suggests the question of whether there is also a physical-science background factor. Since only one field of physical science—physics—was examined in aviation students, this hypothesis could not be tested.

(27) *Planning.* The name of this factor is adapted from the planning tests, in which it is unique. One test of this kind required the examinee to tell in which order he would make the strokes of two successive letters if he were skywriting with an airplane. Another called for the most economical order for a series of errands to be performed in a city whose layout is pictured. This factor might possibly be another kind of visualization, but, unlike the manipulatory visualization, it calls for more of a creative contribution. It is distinct from the judgment factor, although both may occur in the same tests.

As a general consideration, there is always the question of whether factors are completely independent or correlated to some degree. Almost exclusively, orthogonal frames of reference were used in the AAF for describing factor configurations. There seemed to be no evidence of very significant correlations among the factors. The matter has recently been given some attention. There are probably some small correlations among some of the intellectual and perceptual factors. An interesting, slight, negative correlation appeared between the number factor and the mechanical-experience factor. The probable reasons for this are not hard to surmise.

The list of factors is probably longer than many would expect, and yet it is certainly not exhaustive. The fact that other previously reported factors did not emerge should not be taken as evidence against their existence. The absence of a universal or "g" factor in Spearman's sense, however, is a significant outcome.

The list throws into bold relief the limitations of intelligence tests and the IQ. These ordinarily stress most heavily the verbal, general-reasoning, and numerical factors, probably in that order, with sprinkling contributions of space, visualization, and memory, depending upon the test. The IQ, as derived from different tests, is not a constant article by any means. IQ tests would have been almost useless in selecting pilots from among those with IQ's of 100 and above, had mastery of flying training been the criterion. On the other hand, it was estimated that the pilot criterion leaned most heavily upon such resources as spatial relations, mechanical experience, psychomotor coordination, pilot interest, and perceptual speed, with minor dependence upon about 15 of the other factors in the list. Even then, with all these factors optimally weighted in a composite aptitude score, less than 60 per cent of the variance in the pilot criterion would be accounted for. Within the limitations of the reliability of the criterion itself, perhaps as much as 80 per cent could be predicted if all pertinent factors were included in the aptitude composite (*3*). Before the war, a selective test or battery would have been regarded as relatively successful if it accounted for as much as 30 per cent of the variance in a practical criterion.

The vocational implications of all this are quite clear. Fitting individuals to educational and vocational plans presents a much more complicated problem than has heretofore been supposed. The chief hope for a solution appears to lie in the direction of the use of factorial thinking and practice. Persons, tests, and jobs can all be described in terms of the same reference frame—one with which many can agree because it is derived in an operational manner. Not only aptitudes, whether biologically or culturally conditioned, or both, but also temperamental and other personality traits can be translated into the same types of categories. It is probable that substantial gains in the assessment of all aspects of personality await such an approach. It is also probable that controlled fundamental research on other problems of human nature awaits the identification of its real variables.

References

1. CARLSON, H. B. *J. exp. Psychol.*, 1937, 5, 477-492.
2. FLANAGAN, J. C. *Science*, 1947, 105, 57-60.
3. GUILFORD, J. P. (Ed.) *Printed classification tests.* (AAF Aviation Psychology Program Research Reports, No. 5.) Washington: Government Printing Office, 1947.
4. THURSTONE, L. L. *Vectors of mind.* Chicago: University of Chicago Press, 1935.
5. THURSTONE, L. L. *Psychometr. Monogr.*, 1938, No. 1.