

MEASUREMENT OF LEARNING AND  
MENTAL ABILITIES\*

HAROLD GULLIKSEN  
PRINCETON UNIVERSITY  
AND  
EDUCATIONAL TESTING SERVICE

About 27 years ago, a small group of students met with Professor Thurstone in Chicago to discuss methods of encouraging quantitative work in psychology. The initial group that was concerned about the slow rate of development of quantitative work in psychology included Jack Dunlap, Al Kurtz, Marion Richardson, John Stalnaker, G. Frederic Kuder, and Paul Horst. They had discussed the problem, had been helped a bit by Donald Paterson, and had decided that possibly if a magazine were set up to publish quantitative psychological material this would facilitate the development of the field. Persons who did good quantitative work, either theoretical or experimental, would thus have a forum where it would be accepted because it was high quality quantitative work, rather than being rejected because it was quantitative and hence "not of too great interest" to the readers.

It developed after discussion that possibly the best method of supporting such a journal would be to have a society which would have this journal as its major organ. This was the nucleus of the Psychometric Society and of the magazine *Psychometrika*, a quarterly journal devoted to the development of psychology as a quantitative rational science.

Thus, in March of 1936, Volume 1, Number 1 of *Psychometrika* was issued with Marion Richardson as Managing Editor, and Horst and Thurstone as members of the editorial board. From this small beginning with five or ten people interested in furthering the development of the field, it is interesting to look back now and consider what has happened during the intervening 25 years.

Let us look at the state of quantitative rational psychology at that time. Thurstone's work over the preceding ten years, from 1925 to 1935, might well be thought of as typifying the field then. He had done some work in the area of learning (Thurstone [44, 46]), developing certain learning

\*Prepared as a technical report in connection with research partially supported by Office of Naval Research contract Nonr 1858-(15) and National Science Foundation Grant G-3407 to Princeton University, and by the Educational Testing Service. Reproduction of any part of this material is permitted for any purpose of the United States Government.

curves and checking on the fit of these curves to learning data. He had also considered some of the typical material in psychophysics, had become somewhat dissatisfied with the emphasis in psychophysics on measuring brightness of lights or heaviness of weights, had thought that it would be tremendously more fruitful and interesting to measure the strength of an attitude, the beauty of a picture, the degree of preference for a belief, for a nationality, or for a political candidate. This was the genesis of Thurstone's psychophysics—the Law of Comparative Judgment set up to analyze data collected by the experimental method of paired comparisons. Later Thurstone initiated what Torgerson has termed the Law of Categorical Judgment to deal with the data collected by the experimental method of successive intervals. Successive intervals was developed for the situation in which one could not reasonably require that the subject make all intervals equal (method of "equal-appearing intervals") or where there was doubt that he could or would do so, even if requested. At this time also, Thurstone [45] had completed his beginning text on test theory, a photo-offset version, and had started his developments of factor analysis for the further study of mental abilities. Thus he had worked in the various areas which today represent the major areas in which the quantitative rational approach in psychology has achieved the most success.

It is of interest that Professor Boring [5] in a recent discussion of quantitative developments in psychology specified four areas that had been particularly fruitful for such developments. These were psychophysics, learning, mental measurements, and reaction time. Thurstone's work between 1925 and 1935, as indicated above, dealt with three of these four areas.

During the subsequent 25 years there has been relatively little quantitative development in the study of reaction time. There has, however, been a tremendous growth in psychophysics or psychological scaling, in learning, and in mental measurements represented by developments in test theory and in factor analysis. As to the work in psychophysics or psychological scaling, I shall simply refer to the symposium held this morning as an illustration of the development in this field over the last 25 years, and will consider here in some detail Learning, Test Theory, and Factor Analysis.

In order to set the stage for the discussion here I should like to illustrate one view of the relationship between scientific theory, mathematics and statistics (Gulliksen [18]). One always, of course, initially has the psychologically meaningful verbal statements of the postulates, the basic assumptions of any system. The characteristic thing about the mathematical rational approach is that at a very early stage these postulates, that is, the functioning postulates that would have some impact on deducing the nature of experimental results, are translated into the language of mathematics. We then have the stage of mathematical development of the concepts eventuating in various equations some of which contain two or more terms that can be

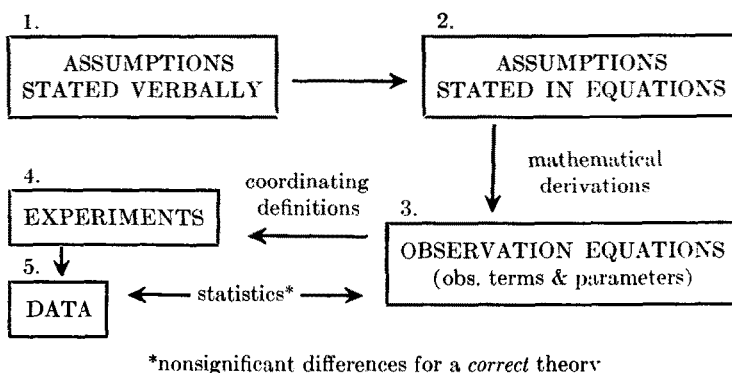


FIGURE 1  
Mathematical Formulation of Psychological Theories

subject to experimental observation. These then may be termed the observation equations for which one can gather data. One then designs an experiment and collects data from the experiment and then (with statistics) checks on the degree of agreement between the observation equation and the data. Frequently when one speaks of quantitative methods in psychology, he is thinking only of the use of statistics to check on the agreement between a hypothesis and data.

In this discussion I will not deal with statistics which is essentially the last step in the development. I will discuss the complex indicated by the verbal psychological statements of the postulates, the mathematical statements of these same postulates, and the derivations from which one gets various implications of the initial postulates eventuating then in mathematical equations that could be in agreement with data from experiments or that could be in disagreement with data.

Statistics (the estimation procedures, testing of hypotheses, and the determination of confidence intervals) is a field that has undergone such tremendous developments in the last 25 years that again it could not possibly be covered even in a symposium devoted entirely to statistics.

Omitting both Psychophysics and Statistics is reminiscent of Sherlock Holmes in "The Adventure of Silver Blaze." When asked for the most significant item in the case to date, he said, "The strange behavior of the dog in the nighttime." Watson, after thinking a moment, replied, "But the dog did nothing in the nighttime." "That," said Holmes, "is the strange behavior."

In the consideration of developments in the last 25 years, in a single symposium, it is necessarily true that the most significant items in development are those that are being omitted because they are too extensive to deal

with short of several symposia. Areas essentially nonexistent 25 years ago are now too extensive to be considered in a single session.

### *Learning*

One area in which there has been considerable development of mathematical translation of verbal postulates and derivation of their consequences is the area of learning (Hilgard [21]). Thurstone [44, 46] in the early 30's developed a theory based on an analogy of sampling from an urn, and showed that the equations derived from such assumptions were in reasonable agreement with data. Since then there have been a number of learning theories stated in mathematical form. Gulliksen [16, 17] has generalized Thurstone's initial equations and developed others based directly on Thorndike's law of effect showing that these equations are identical with those that Thurstone developed in terms of an urn model. Rashevsky [35] has taken an approach from basic ideas of the functioning of the nervous system, utilizing inhibition and facilitation, and has developed some equations of learning on this basis.

Hull [24] has utilized as his starting point the conditioning model where each repetition has an effect of increasing the strength of the response. He also used the concept of confusability of various responses to account for the lack of, shall we say, immediate learning to explain different degrees of difficulty of learning in serial lists. The probabilistic model that expresses its postulates in terms of operators increasing and decreasing the probabilities of response has also been developed during this time (Bush and Mosteller [7]).

I should also mention the work of Audley [3] in London. He has developed probabilistic equations of learning and devised methods of fitting these to individual learning curves so that one can obtain parameters for each individual from data on learning curves and also from data on changes in reaction time with learning. This is a rather interesting development, first, because it develops the probabilistic model so that parameters can be computed for each individual, and, second, because it relates the right-wrong response data to the reaction time data. One of the characteristics of learning is that the reaction time usually decreases. This theory tries to show that these two curves are two different manifestations of the same basic set of parameters. Roger Shepard [39] has related work in learning to psychophysics showing that generalization in learning is related to psychological similarity.

There have also been some recent interesting attempts to develop these models of learning and to express them in terms of electronic computing machine programs where the machine is instructed to compute probabilities in accordance with the numbers in certain cells. Under reward conditions it adds something to the numbers in those cells, under punishment conditions it subtracts something. The information processing language (described by Green [14]) developed by Newell, Shaw, and Simon is an illustration of this particular approach (see also Newell and Simon [34]). Also Block, Rosenblatt,

and others at Cornell have been working on the perceptron (see Rosenblatt [37]). This is a mechanical gadget in which the initial connections are purely random. However, there is a programming of an increase and decrease in resistance of certain circuits corresponding to reward and punishment and it turns out that this machine with purely random connections is capable of learning. Other discussions of complex behavior of computers are found in the Western Joint Computer Conference proceedings [53], the Teddington National Physical Laboratory symposium [33], Hagensiek [20], Shannon and McCarthy [38], and Uhr [51].

A very interesting thing to note as one surveys these various theories by Audley, Estes, Bush, Mosteller, Hull, Rashevsky, Thorndike, Gulliksen, and Thurstone is the essential similarity in the basic framework of each theory. This can be indicated as follows.

1. There is some procedure to effect the "stamping in," the "facilitation," or the "increase in probability" of a response that in some sense is a correct response, a rewarded response, or a response that is at least dominantly rewarded.

2. There is a corresponding postulate regarding the "stamping out," "inhibition," or "decrease in probability" of a response that may be thought of as a wrong response, an incorrect response, an unrewarded response, or at least a dominantly nonrewarded response.

3. Many of the theories also have some provision regarding resemblance or similarity of stimuli either in their sensory characteristics or in their position, such as position near to each other in a rote learning series. This sort of similarity or contiguity leads in certain contexts to confusion and slows up learning; in other contexts it is termed "generalization of response to similar stimuli," or "transfer of training," or "equivalence of stimuli." Some mechanism, in other words, whereby a response which has initially been learned to one stimulus tends to be given to other stimuli. Depending on the particular learning set-up designed by the experimenter this tendency may either delay learning in one situation, or facilitate generalization in another situation.

4. There is also some sort of decrease in probability or fading out of a response, "forgetting" due either to passage of time or due to confusion with other stimuli. In some guises it has been termed retroactive inhibition. Rashevsky has shown how a differential decline rate for inhibition and facilitation could produce a "reminiscence" effect. This decline with time again enters into a number of the different learning theories.

5. There is also a change in reaction time that is often made a part of the theory. Hull utilized this as one of his postulates. One of the manifestations of learning is a decrease in response latency. Audley [3] has also used this to give a very interesting possibility for a sort of reliability check on a single learning situation.

During the last 25 years we have had a reasonable proliferation of slight variants on the increase or decrease of strengths and probabilities. These various sets of postulates result in somewhat different observation equations. However, the basic observation equations would all be in a superficial sense fairly similar so that it would probably take rather a precise test of a fit in various experiments to determine that one of these theories was a better fit to the data than others. Bush and his co-workers at Pennsylvania are embarking on such a program now. It is to be hoped that others will follow and that in the next 25 years we will be able to specify more accurately the kind of learning situation for which a given model or equation is most appropriate.

### *Reliability of Learning Parameters*

I want to mention here a development that is not, strictly speaking, quantitative, but one that may have a tremendous influence in the quantitative development and testing of learning theory. This stems from the work of Sperry [40]. He has found it possible to divide a brain into two halves by sectioning the corpus callosum and the optic chiasma; he reports that not only is it found that habits learned by one half do not transfer to the other half, but for a given animal the peculiarities manifested by him in "right brain learning" are again exhibited in "left brain learning." Should this turn out to be verified, or generally true, we now have a possibility never before envisaged by workers in the field of learning, the "split brain reliability."

In my opinion one of the great handicaps under which work in learning has labored over the last hundred years has been the fact that unlike the mental test area, it has been essentially impossible to do a repeat experiment and to determine a reliability. Every respectable achievement or aptitude test has some device of odd-even, first and second half, or repeat test, whereby one attempts to do the same thing twice and measures the accuracy of the technique by the correlation between these two halves—the reliability coefficient. In the case of learning the experimenter could always obtain a learning curve to determine parameters. However, when he attempted to get another learning curve, there was always a dilemma. He could experiment on animals which had not been used for the first set of learning curves, in which case there was simply a sort of species reliability. It would be considered extremely poor procedure, in the case of an intelligence test, to correlate one person's score with another person's score in order to determine the test reliability. Or he could have the same subjects learn another problem, in which case there was always the question, "Was the subject learning the second problem better because of the influence of the first one, or was he hindered in his learning of the second problem because of the influence of the first one?" The experimenter could never be particularly certain which was the case and, as a result, measures of learning have not had reliability coefficients attached. One just does not know the extent to

which the lack of agreement is a result of a difference in the psychological function being tested, the psychological ability being tested, or simply the result of poor experimental techniques. Certainly this contribution of Sperry and others is worth an extremely careful look to see if the initial possibility that it holds for "split brain reliability" coefficients in the case of learning tasks is really borne out.

### *Relation of Intelligence to Learning*

I should also like to emphasize that while one purpose of learning theories is, of course, to describe the course of learning, this in itself should not stand as a final goal. Important questions can be raised regarding the relationship of these learning parameters to other parameters characterizing behavior of the individual. I can illustrate this point with the studies by Stake [41] and Allison [1]. They both have raised a question regarding the relationship between mental abilities and learning. As we know, for decades intelligence has been defined as the ability to learn, yet intelligence tests have measured the ability to learn not directly but only by inference. They have concentrated on what has already been learned. Both Stake and Allison have set up a variety of learning problems, have fitted equations of the learning curve to the data obtained from 200 or 300 persons who took these learning tests, have also given these people some 30 or 40 aptitude and achievement tests and then have entered the entire material into a factor study. The purpose of these studies is to determine how many different learning abilities there are, and to see how these learning abilities are related to the abilities measured by aptitude and achievement tests.

First we can say that, as a result of these two studies, the learning area is definitely a complex area that cannot be represented in terms of one learning ability. There are many different kinds of learning ability—how many we will not know until a good many more studies have been made. Second, it is clear that some of the abilities required for the learning tasks are *not* represented in any of the intelligence measures. The nature and the importance of these abilities that have been missed by the one-shot aptitude and achievement measures constitutes a very important problem for further investigation.

I should also indicate that studies such as Stake's and Allison's could not have been conducted without electronic computers. Stake estimated that by Monroe-Marchant methods in use a few years ago his analysis would have taken one hundred and twelve man-years. With electronic computers the job was done in about six months.

### *Master- or Reference-Learning Curves*

In the first volume of *Psychometrika*, Eckart and Young [11] published a very important paper. It dealt with the approximation of one matrix by

another of lower rank. It applied in general to any matrix, square or rectangular, and furnished the essential basis for the use of matrix theory for expressing and testing a large number of quite different psychological hypotheses. (See also Hohn [22] for an elementary treatment of matrices.)

One interesting application of the Eckart-Young theorem is to learning matrices. For many years, people have analyzed group learning data, plotted group learning curves, and criticized others on the ground that there are individual differences in learning which averages ignore. The Eckart-Young procedure has been used by Tucker [48, 50] for analyzing learning data. The matrix of trials by individuals is factored to give a minimum number of "reference" or "master" learning curves. Each individual receives a set of weights indicating the extent to which he has utilized each curve. If the matrix is rank one, then there is only one master learning curve, and the average curve is a good representation for each individual. In general, for ranks greater than one the individuals will not be correctly represented by the average curve.

Tucker [48, 50] has applied this method of handling learning matrices to some probability learning data collected by R. Allen Gardner. He finds that in a simple probability learning situation where the subject is distinguishing between probabilities of .70 and .30, the matrix is of rank one. Only one learning curve is necessary to explain the data. In another situation, where four objects were presented with relative frequencies 70, 10, 10, and 10 percent, three different learning curves were needed to explain the data. There were apparently (shall we say) early learners, medium learners, and people who caught on to some of the ideas very late in the series of trials, so that one of the learning curves was a rapidly rising negatively accelerated curve, and the other two were inflected *S*-shaped curves. The different subjects had different weighted combinations of these curves.

Weitzman [52] has utilized the Eckart-Young procedure for analyzing matrices of learning data (animals by trial matrices) for a combined group of rats and a group of fish, putting them together as successive rows of the same matrix and applying a uniform analysis. The question is, "Will the learning curves that are necessary for the rats be the same as those that are exhibited by the fish, and will the weights of the learning curves needed for the rats be the same as or different from the weights needed for the fish?" In his particular case he found a rather clear-cut rank-two structure which means that the same two, shall we say, master learning curves were necessary to explain the learning data for the rats and for the fish.

### *Test Theory*

The area of mental measurement, which in the 30's was represented by Thurstone's [45] small photo-offset manual, now covers a huge literature (Anastasi [2], Cronbach [9], Guilford [15], Thorndike and Hagen [43], Lindquist [27], Remmers and Gage [36], and Meehl [31]).



Reliability and error of measurement are no longer the simple concepts they were 25 years ago (Cureton [10], Jackson and Ferguson [25]). Guttman [19] has developed formulas for lower bounds of reliability coefficients. Cronbach [8] has suggested many different kinds of reliability coefficients taking account of various types and combinations of factors which can affect test performance. Perhaps one generalization would be to point out that there are  $k$  different factors which may influence test performance such as fatigue, practice, additional learning, time of day, state of health, emotions, distractions, maturation, and growth. There are then  $2^k$  different reliability coefficients, depending on which particular set of factors is of interest for the particular use to be made of the test. The more important ones have been explicitly dealt with by Cronbach, Guttman, and others.

Error of measurement is no longer a single number to attach to a test to represent variance of observed test scores for persons with the same true score. The error of measurement is a function of true score, so that the discriminating power of the test will be different at different ability levels. Mollenkopf [32] initiated some work in this area. The problem is being studied in greater detail by Birnbaum [4] and Lord [30]. The goal of this work would be to develop procedures so that it would be possible to specify the discriminating power desired in various ability ranges, and then to construct a test having the desired characteristics.

The personnel classification problem is the problem of assigning or recommending the most efficient utilization of each person in a group to perform the set of jobs to be done by that group. Votaw, Brogden [6], and others have suggested solutions for the problem.

The central problem of test theory is the relation between the *ability* of the individual and his *observed score* on the test. A third concept, that of the true score of an individual on a test, has also been introduced in an effort to clarify the problem. Psychologists are essentially in the position of Plato's dwellers in the cave. They can know ability levels only through the shadows (the observed test scores) cast on the wall at the back of the cave. The problem is how to make most effective use of these shadows (the observed test scores) in order to determine the nature of reality (ability) which we can know only through these shadows. Birnbaum [4], with his studies of test theory, and Lazarsfeld [26], with his use of various trace lines in latent structure analysis, have proposed various types of solutions to this problem.

An attempt to develop a consistent theory tying test scores to the abilities measured is typified by Lord's recent work [28], including his Psychometric Society presidential address [30], in which he formulated at least five different theories of the relationship between test scores and abilities, and showed how it was possible to test certain ones of these. It is to be hoped that during the next 10 or 20 years a number of these tests will be carried out so that we will have not five different theories of the relationship between ability and test score and various possible trace lines, but we will be able to

say that, for certain specified tests constructed in this way, here is the relationship between the score and the ability measured, and this is the appropriate trace line to use.

### *Factor Analysis*

Another one of the major developments over the last 25 years has stemmed from the work in factor analysis of mental tests. It is interesting to note that when Thurstone worked for the military during the first world war, the contribution of psychologists under Dr. Yerkes was to set up a single measure of ability, the Army Alpha, or a measure of lower level ability, the Army Beta, and to range all men along the single scale of the Army Alpha test and on the strength of this information to assign jobs.

I remember in teaching beginning psychology classes in the late 20's that I repeatedly explained to doubting freshmen that it was merely a popular superstition that some people had high verbal ability and others had high mathematical ability. These various abilities were perhaps matters of differential interest, but basically there was only one intelligence as indicated by the Spearman so-called two-factor theory, which of course was one general factor with various sorts of specific factors, and that any belief in various factors had the status purely of an unverified popular superstition.

In the early 30's Thurstone took the view that very possibly we had failed to find different types of intelligence simply because we had not looked carefully enough with sufficiently powerful methods. He developed the factor methods, found that there was a mathematics—the mathematics of matrix theory—that was possibly relevant, and devoted his time to studying this and applying it in the analysis of mental abilities. I remember Thurstone telling that he had presented his factor problem to some of the mathematicians at a Quadrangle Club lunch one noon, pointing out that he had a square array of numbers here (the set of correlation coefficients), that he wanted to get one rectangular array such that when multiplied together in a certain way the sum products of the numbers in these two rectangular arrays would equal the correlations in the one larger square array. He said they smiled at each other and said, "Oh, the square root of a matrix is all that is." He insisted on pursuing the inquiry further, found that there was a field that possibly dealt with this topic that he should be interested in, tutored in it for some years, and developed as a result the vectors of mind and multiple factor analysis. Tremendous numbers of studies stemmed from this work. Other theoretical developments in the area were made by Truman Kelley and Harold Hotelling, who also generalized Spearman's one general-factor view to include the possibility of a large number of factors. This was the beginning of literally hundreds of factor studies which led to the development of a variety of tests of various mental abilities. One illustration of the impact of this work is the difference in the testing program in the second

world war. None of the services utilized only a single measure of general intelligence. There were tests of a variety of abilities—verbal, quantitative, spatial, mechanical. Placement for different types of assignments was dependent on different weighted combinations of these abilities.

### *Theory of Factor Analysis*

With respect to the theoretical developments in factor analysis, we have had a considerable growth in the area of statistical tests for significance of factors or of ranks of matrices, although considerable still remains to be done in this area. The development of methods of comparing factor analyses results of one battery with those of another—the interbattery method—constitutes an extremely significant contribution (Tucker [49]). The other lack, until recently, was the lack of methods for comparing one study on a given set of tests with another study using the *same* set of tests on a *different* sample of people (Tucker [47]). So we now have precise methods for comparing different groups given the same battery, and different batteries given to the same group. These are powerful extensions of the factor method.

The recent development of high-speed computing methods is also critical for this field. Twenty years ago there was a considerable argument between persons with a mathematical bent, such as Hotelling, who insisted that one must use the principal axis solution, and experimenters, such as Thurstone, who maintained that, while the principal axis solution was very nice, he had never seen anyone utilize it with 50 tests on 200 or 300 people. We now of course have computing routines that give the principal axis solution at a feasible time and cost so that this controversy is now technologically obsolete. Thurstone would clearly have adopted the principal axis solution as soon as it was feasible from the point of view of cost involved and time consumed.

Many of the problems in test theory and factor analysis are essentially problems of multivariate analysis in mathematical statistics. It is very encouraging to note that many psychologists are developing proficiency in mathematical statistics, and also that mathematical statisticians, such as T. W. Anderson, Frederick Mosteller, David Votaw, Allan Birnbaum, D. N. Lawley, M. G. Kendall, S. S. Wilks, John Tukey, and others, are becoming interested in some of the statistical problems associated with test theory and other branches of psychology, and are providing the psychologists with solutions to these problems.

### *Applications of Factor Analysis*

There have been various conferences on factor analysis and its results lately. Two monographs by French [12, 13] on the various achievement and aptitude factors and the various personality factors indicate the degree to which this field has proliferated. The need now seems to be for more systema-

tization, boiling down, determining which of the factors are important and which are not, rather than added proliferation of the factors.

Typically, the work in factor analysis has dealt with a battery of *predictors*. However, increasing attention is being directed toward the problem of using a battery for efficient prediction, differential prediction, of multiple criteria. The Psychological Corporation has a differential prediction battery. Horst [23] at the University of Washington has been developing the theory for differential prediction, and developing such a battery.

#### *Achievement Tests*

I probably should also mention that the field of achievement testing has developed considerably since the early 1900's, when three-hour essay examination graded by crews of readers was the standard procedure for the College Entrance Examination Board. There is some appreciation of the fact that evaluation of the essay is not very precise, and that teachers need to be taught the appropriate methods for preparing and evaluating classroom tests. This is an extremely large job on which only a relatively small start has been made as of now. In the next 25 years I would hope for considerably greater sophistication of the classroom teacher in the development and evaluation of tests than we find now.

#### *Summary*

We have considered developments over the last 25 years in the area of measurement of mental abilities. Marked advances have been made in determining the relationship between the ability measured and the test score, in methods of item analysis, in the differentiation and classification of various methods of dealing with reliability. The big development in this area though has been the change from the emphasis on a single general intelligence to the differentiation of a large number of different aptitudes. This has been made possible by the development of the factor analysis methods.

Note that factor methods were just at their beginning when *Psychometrika* was started, that the initial paper by Young and Householder on multi-dimensional scaling techniques had not yet been written, the Eckart-Young paper dealing with the expression of one matrix as a product of two other matrices of minimum rank, a fundamental factor analysis theorem, had not yet been written, and that the factor computations were done entirely with Monroe-Marchant methods. We can see that during the last 25 years there has been, first, a terrific growth in the basic theory related to mathematical formulation of psychological problems—basic theory in the area of testing, in the area of aptitude measurement and factor analysis, in the area of learning, and in the area of psychophysics. Second, there has been a tremendous development of computational methods, enabling us to do studies now that

were essentially impossible because of time and cost factors even five or ten years ago.

The findings resulting from these methods have an impact in various areas. The development of multiple factor tests has changed the entire picture of the testing field from what it was during the first world war. The development of a variety of learning theories gives some promise that in the next 25 years we will be able to specify the types of conditions, if any, under which these various theoretical approaches are appropriate.

The development of the unidimensional and multidimensional scaling methods and their use in a variety of areas, in measuring sensations, in measuring preferences or values for objects, should have considerable impact. Various fields such as linguistics, sociology, and economics should benefit tremendously from some of these methods that have been developed during the last 25 years since this small group of students met with Thurstone and decided to form the Psychometric Society to publish *Psychometrika*, and to further the development of psychology as a quantitative rational science.

#### REFERENCES

- [1] Allison, R. B., Jr. Learning parameters and human abilities. Princeton Univ., Psychol. Dept., 1960. (multilith)
- [2] Anastasi, A. *Psychological testing*. New York: Macmillan, 1954.
- [3] Audley, R. The inclusion of response times within a stochastic description of the learning behavior of individual subjects. *Psychometrika*, 1958, **23**, 25-31.
- [4] Birnbaum, A. Probability and statistics in item analysis and classification problems: on the estimation of mental ability. Ser. Rep. 15. Pensacola, Florida: Sch. of Aviation Medicine USAF, 1957.
- [5] Boring, E. G. The beginning and growth of measurement in psychology. Paper presented at the N. R. C. and S. S. R. C. conference on "The history of quantification in the sciences," Nov., 1959.
- [6] Brogden, H. E. Least squares estimates and optimal classification. *Psychometrika*, 1955, **20**, 249-252.
- [7] Bush, R. and Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.
- [8] Cronbach, L. J. Test "reliability": its meaning and determination. *Psychometrika*, 1947, **12**, 1-16.
- [9] Cronbach, L. J. *Essentials of psychological testing*. New York: Harper, 1949.
- [10] Cureton, E. E. The definition and estimation of test reliability. *Educ. psychol. Measmt*, 1958, **18**, 715-738.
- [11] Eckart, C. and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, **1**, 211-218.
- [12] French, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monogr. No. 5*. Chicago: Univ. Chicago Press, 1951.
- [13] French, J. W. The description of personality measurements in terms of rotated factors. Princeton: Educ. Testing Serv., 1953.
- [14] Green, B. F., Jr. IPL-V: The Newell-Shaw-Simon programming language. *Behav. Sci.*, 1960, **5**, 94-98.
- [15] Guilford, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.
- [16] Gulliksen, H. Rational equation of the learning curve based on Thorndike's law of effect. *J. gen. Psychol.*, 1934, **11**, 395-434.

- [17] Gulliksen, H. A generalization of Thurstone's learning function. *Psychometrika*, 1953, **18**, 297-307.
- [18] Gulliksen, H. Mathematical solutions for psychological problems. *Amer. Scientist*, 1959, **47**, 178-201.
- [19] Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, **10**, 255-282.
- [20] Hagensick, P. W. Logic by machine: programming the LGP-30 to solve problems in symbolic logic. *Behav. Sci.*, 1960, **5**, 87-94.
- [21] Hilgard, E. R. *Theories of learning*. (2nd ed.) New York: Appleton-Century-Crofts, 1956.
- [22] Hohn, F. E. *Elementary matrix algebra*. New York: Macmillan, 1958.
- [23] Horst, P. and MacEwan, C. Optimal test length for multiple prediction: the general case. *Psychometrika*, 1957, **22**, 311-324.
- [24] Hull, C. L. *Principles of behavior*. New York: Appleton Century, 1943.
- [25] Jackson, R. and Ferguson, G. Studies on the reliability of tests. Toronto: Dept. Educ. Res. Bull. No. 12, 1941.
- [26] Lazarsfeld, P. F. Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. III. New York: McGraw-Hill, 1959.
- [27] Lindquist, E. F. (Ed.) *Educational measurement*. Menasha, Wis.: Banta, 1951.
- [28] Lord, F. M. A theory of test scores. *Psychometric Monogr. No. 7*. Chicago: Univ. Chicago Press, 1952.
- [29] Lord, F. M. Tests of the same length do have the same standard error of measurement. *Educ. psychol. Measmt*, 1959, **19**, 233-239.
- [30] Lord, F. M. An approach to mental test theory. *Psychometrika*, 1959, **24**, 283-302.
- [31] Meehl, P. E. *Clinical and statistical prediction*. Minneapolis: Univ. Minnesota Press, 1955.
- [32] Mollenkopf, W. G. Variation of the standard error of measurement. *Psychometrika*, 1949, **14**, 189-230.
- [33] National Physical Laboratory, Teddington, England. *Mechanisation of thought processes*. London: Her Majesty's Stationery Office, 1959.
- [34] Newell, A. and Simon, H. A. The simulation of human thought. Santa Monica, Calif.: The RAND Corp., 1959. (mimeo)
- [35] Rashevsky, N. *Mathematical biophysics*. (Rev. ed.) Chicago: Univ. Chicago Press, 1948.
- [36] Remmers, H. H. and Gage, N. L. *Educational measurement and evaluation*. New York: Harper, 1955.
- [37] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, 1958, **65**, 386-408.
- [38] Shannon, C. E. and McCarthy, J. (Eds.) *Automata studies*. Princeton: Princeton Univ. Press, 1956.
- [39] Shepard, R. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55**, 509-523.
- [40] Sperry, R. W. Physiological plasticity and brain circuit theory. In H. F. Harlow and C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison, Wis.: Univ. Wisconsin Press, 1958. Pp. 401-424.
- [41] Stake, R. Learning parameters, aptitudes and achievements. Princeton Univ., Psychol. Dept., 1958. (multilith)
- [42] Thorndike, E. L. *The fundamentals of learning*. New York: Columbia Univ. Bur. Publ., 1932.
- [43] Thorndike, R. L. and Hagen, E. *Measurement and evaluation in psychology and education*. New York: Wiley, 1955.
- [44] Thurstone, L. L. The learning function. *J. gen. Psychol.*, 1930, **3**, 469-493.

- [45] Thurstone, L. L. *The reliability and validity of tests*. Ann Arbor, Mich.: Edwards Brothers, 1931.
- [46] Thurstone, L. L. The error function in maze learning. *J. gen. Psychol.*, 1933, **9**, 288-301.
- [47] Tucker, L. R. A method for synthesis of factor analysis studies. Dept. of the Army, Personnel Res. Br., AGO, Rep. 984, 1951.
- [48] Tucker, L. R. Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 1958, **23**, 19-23.
- [49] Tucker, L. R. An interbattery method of factor analysis. *Psychometrika*, 1958, **23**, 111-136.
- [50] Tucker, L. R. Determination of generalized learning curves by factor analysis. Princeton Univ., Psychol. Dept., 1960. (multilith)
- [51] Uhr, L. Intelligence in computers: the psychology of perception in people and in machines. *Behav. Sci.*, 1960, **5**, 177-182.
- [52] Weitzman, R. A. A comparison of the performance of rats and fish on a probabilistic, discriminative learning problem. Princeton: Educ. Testing Serv. Rep., 1959.
- [53] Western Joint Computer Conference. *Proceedings of the 1955 Western Joint Computer Conference*. New York: Amer. Inst. of Elect. Engineers, 1955.