COMMUNICATIONS AND DISCUSSIONS

SUGGESTIONS WITH REGARD TO PROFESSOR THURSTONE'S "METHOD OF CRITICAL SCORES"

THE writer was much interested in Professor Thurstone's vigorous demand, in a recent article in this Journal¹ for a greater consideration of the practical problems of individual diagnosis, in evaluating mental tests. The few comments below are intended not as criticism of Professor Thurstone's methods, but rather as an effort to carry his general plans still further than he has done.

In the first place, if a group of tests is to be used for the purpose of prognosticating failure in college why not evaluate the tests directly according to their ability to do this, -- instead of by the more round-a-bout comparison of their agreement with instructors estimates of ability? Various rough and ready methods for measuring the "efficiency" of a test in differentiating the extremes of a group are possible. The writer has recently been working over some results from a large sixth grade class in Junior High School, with a brief group scale of intelligence, in order to compare the differential value of the tests of this scale in distinguishing those pupils who failed at the end of the school year. The method used is extremely simple and straightforward. The results for the sixth grade on each test had been already tabulated. At the end of the year the writer simply obtained from the school a list of the failures and marked the score of each one of these children over in red ink. This, in itself, gave a rough picture of the situation; the extent to which the red marks of the failures were found well up among the pencil marks indicating the scores of those who passed in their work for the year showed the extent to which the particular test in question failed to differentiate the failures from the successful children. A numerical statement of differential power was obtained simply by counting the number of pencil marks below each red mark, -- that is, the number of successful children which the test failed to distinguish from each particular failure -- and averaging the results for all the red marks, for each test. This final average gave the "average error" which, expressed in percent, gave a "percent of error" for each test. The method has been described more fully elsewhere², and need not be further detailed here.³ As illustrative

¹ Thurstone, L. L. Tests for College Freshmen, *Journal of Educational Psychology*. Vol. 10. pp. 129-92, March 1919.

² Pressey, S. L. and L. W. The Practical "Efficiency" of a Group Scale of Intelligence, *Journal of Applied Psychology*. Vol. 2, pp. 68-80, March 1919. A preliminary description of the "Cross-out" scale, from study of which the illustrative material of the present brief paper has been drawn, appeared in the June number of the same journal; a final statement will be published shortly.

³ The further point should be mentioned, however, that although the method weights, heavily, marked discrepancies between test finding and the independent criterion it does not, as does the "method of critical scores" put the situation altogether at the mercy of the single extreme deviate score.

of the general run of the figures to be expected it may be said shortly that the writer's tests have shown from four to twelve per cent error in differentiating failures.

(2) A second suggestion has to do with the form that the recommendation to the school authorities, based on the test findings, should take. Professor Thurstone tells us that "Those students are reported as doubtful who attain a median percentile rank of 22% or below, and who are below the lower critical score on two or more of the six tests" (p. 136). But an administrative officer would surely be interested to know the degree of doubtfulness implied in this statement. He might also wish to know what inference might be drawn from a score at the five percentile or at the twenty-five percentile. A method which gives an exact statement of the probability of failure implied by a score at any point in the distribution may be illustrated by further treatment of the results from the "Cross-Out" scale above mentioned; results are again from the sixth grade, (156 cases). The tests were given at the beginning of a semester; at the end a list of the failures was obtained, as stated above, and also a list of children who were considered by at least three out of the four teachers in the four promotion subjects, to be doing brilliant work. There were 14 children doing exceptional work and 13 complete failures (children who failed in four or more subjects). The scores of the failures were marked in on the distribution of the total scores for the entire grade in red ink; the scores of the brilliant children in black ink. The deciles for the whole distribution were now calculated, and the number of failures and of brilliant children scoring in each tenth of the distribution noted and expressed as a percent of the total number in the division. The final result was the following table of probabilities:

TABLE I

Decile divisions	1	2	3	4	5	6	7	8	9	10
Percentage Probability that case will fail	47	13	20	7	7	0	0	0	0	0
Percentage probability that case will do brilliant work	0	0	0	0	0	0	7	20	20	47

For use in individual diagnosis such a table should be read: if a given child in the sixth grade, at the first of the semester, scores in the lowest tenth of his class, there is a 47% probability that he will be a complete failure at the end of the year; if he scores in the second decile division, there is a 13% probability etc. for the other divisions. The second line is read similarly for prognosis of brilliant work. The above figures are of no great importance as they deal with prognosis based on a relatively few cases, over a short period of time, but the method would appear to have some general application.

(3) Professor Thurstone very correctly points out the inadequacy of the correlation coefficient in giving information with regard to the ability of a test to differentiate extreme cases. However, the problems of individual diagnosis may also arise with regard to cases around the average in ability, -- as in making up "fast," "medium" and "slow" sections in Junior High School. A way in which a scatter diagram may be thrown into a simple form readily interpretable by psychological laymen, for purposes of individual prognosis in such situations, is given below. The basis of the calculations was a correlation table for the sixth grade, showing the relationship between test scores and pooled ratings of each pupil's general ability obtained, by use of a rating scale, from four teachers. The ratings were made in June after the teachers had had a year of acquaintance with the children. This table was divided into 25 parts by drawing lines through the 20, 40, 60, and 80 percentiles on both distributions; the table was thus divided each way into five sections, not equal in size on the plotting paper, but equal in having the same number of cases. It was then summarized to read for 100 cases, twenty in each of the five vertical, and twenty in each of the five horizontal divisions. The results are given in table II; 1 is high and 5 low in each distribution.

TABLE II

Division into fifths according to test scores	Division into fifths according to teachers' estimates:—				
	5	4	3	2	1
1	0	1	1	6	12
2	2	3	4	5	6
3	3	5	6	5	1
4	5	5	6	3	1
5	10	6	3	1	0

The table is to be read as follows (supposing the scale to have been given to a sixth grade at the beginning of a semester for the purpose of obtaining a prognostication as to ability each child will have shown at the end of the semester). If a given child scores below the 20 percentile, the probability is 10 out of 20 that he will be considered by the teachers, after a year's acquaintance, to belong in the poorest fifth of the class as regards ability; there are 6 chances out of 20 that he will be considered in the second fifth; there is only one chance out of 20 that he will be considered above the third, or average, group. If the pupil tests as in the middle fifth of the class, the chances are 3 out of 20 that he will be judged in the poorest fifth, and one out of 20 that he will be judged in the best fifth, -- and so on.

As methods of presenting test data for use in individual diagnosis these two tables have the practical advantages that (1) they are readily understood, (2) the practical implication of the test findings is direct and unmistakeable, (3) an exact statement of the definiteness of that implication, for any case, is given.

SIDNEY L. PRESSEY

Department of Psychology, Indiana University.

*It has been outlined in a somewhat different way in a previous article. Irregularity on a Psychological Examination as a Measure of Mental Deterioration: *Journal of Abnormal Psychology*. December, 1918.

THE RELATION OF SCHOLARSHIPS TO THE YERKES AND TERMAN ADULT TESTS

During the present collegiate year, students taking courses in Psychology and Education were tested by both the Yerkes-Rossy* adolescent-adult point scale and the Terman adult tests, by the group method. No freshmen are admitted to these classes. The results have been tabulated for the other three classes separately and for the Juniors and Seniors combined as the two latter classes take very nearly the same collegiate work. An Extension class of teachers, the majority of whom had taken 2-4 years of college work and in addition had much experience in teaching, were also given these tests. Two small classes of high school Juniors and Seniors took the Yerkes-Rossy tests. Thus there was here opportunity to compare scholastic performance with intelligence rating by two scales and also note possible diagnostic values of the different exercises.

The following tables give the mean and mean variation of each of these groups for each of the twenty exercises of the Yerkes-Rossy scale.

*Yerkes and Rossy, A Point Scale For the Measurement of Intelligence, Boston Medical and Surgical Journal CLXXVII, 564-573.