

12. PERFIELD, E. E. What Is Rhythm. *Musician*, 1917, 22, 735.
13. PERFIELD, E. E. Exercises for Developing Rhythm and Note Values without Fractional Reasoning. *Musician*, 1917, 22, 94, 252-253.
14. RUCKMICH, C. A. Visual Rhythm. *Studies in Psychology: Titchener Commemorative Volume*, Worcester, Mass.: Louis N. Wilson, 1917, 231-254.
15. RUCKMICH, C. A. A Bibliography of Rhythm. *Amer. J. of Psychol.*, 1918, 29, 214-218.
16. SEASHORE, C. E., & MOUNT, G. H. Correlation of Factors in Musical Talent and Training. *University of Iowa Studies in Psychology, No. VII, Psychol. Monog.*, 1918, 25, 47-92.
17. WEIDENSALL, J. *Studies in Rhythm*. Cincinnati: Bohnett, 1916. Pp. 40.

CORRELATION

BY JAMES BURT MINER

Carnegie Institute of Technology

The Evaluation of Tests.—The problem of measuring the diagnostic value of tests by correlation has come very clearly to the front through the contributions of Ruml. Following his paper of last year on the measurement of the efficiency of mental tests he has deduced what he terms the "rank-tangential coefficient" (33). He proposes using the new coefficient for measuring the value of tests for selecting any portion of a group in order of rank. As a corollary to this, the rank-tangential coefficient gives a measure for determining what portion of a group can best be selected by a given test. This measure may often be more important in practical situations than the correlation coefficient which is the basis of determining only the slant of the smoothed regression line. Ruml shows from the results of tests on three different groups of college freshmen that the tests may be decidedly more valuable for selecting, within the class, groups of similar mental ability than for predicting relative standing within these groups. For two freshmen classes the series of tests which showed a product-moment correlation of about .65 with teachers' estimates gave rank-tangential coefficients of over .90 for selecting the lowest 10 per cent. of the freshmen.

The rank-tangential coefficient is deduced from Pearson's bi-serial r , which Ruml suggests calling a tangential coefficient (T) instead of a correlation coefficient, since it is equal to r only under certain conditions, although it is equivalent in meaning and varies from $+1$ to -1 . The rank-tangential coefficient, " t ," is easily computed and expresses "the ratio of the mean value of X devia-

tions in rank associated with a Y array, to the mean deviation of that Y array." In using the rank-tangential coefficient, it should be remembered that units of rank are treated as equal and the measure of relationship is for resemblance in ranks rather than for resemblance in other measures of ability.

In another paper, Ruml (32) raises a second important objection to regarding the correlation coefficient as always the best indication of the diagnostic value of a test. He plots the relation between the coefficient of correlation and the standard error of prediction of x from y . The increase in the error of prediction is very rapid with the decrease in the size of the coefficient; so rapid, in fact, that when a correlation is as low as .50 the error of prediction is 7/8 of its maximum size. Instead of using r , Ruml, therefore, proposes using the error of estimate as a measure of the diagnostic value of a test.

Rosenow (30), for the first time in psychological work, calls attention to an important characteristic of the multiple correlation coefficient, which suggests a minimum value below which it may have little value even when it is large compared with its own error. He uses Yule's approximation formula to discover how large the coefficient might be owing to the fluctuations of sampling among wholly uncorrelated variables, on account of the fact that the errors do not neutralize each other but are cumulative. Thus, in his study of the multiple correlation between the class standing and 15 tests from 92 cases a multiple correlation coefficient of .40 might arise from variables which were not related at all. The actual coefficient found, .55, when all 15 tests were properly weighted, therefore had little value. On the other hand, with the five best tests a multiple coefficient of .52 is quite significant compared with the value .21 which might be obtained by chance.

The theory underlying Hardwick's method (14) for a new approximation to the best weighting in points to be given for each of the 20 tests in the Yerkes-Bridges-Hardwick Point Scale is hardly justified statistically. It is, however, perhaps better than the common practice of giving a person a summated score from a series of tests without weighting them at all. She assumes that "the tests would be correctly weighted if their scores were proportioned to their respective correlations with general intelligence." She then proceeds to find the average correlation for each test for three groups, 6-year-olds, 9-year-olds, and 12- and 13-year-olds between that test and the total point score for the individuals. The best weighting could only be obtained by partial correlation.

It is essential that allowance be made for the intercorrelations of the tests. Giving the same individual time after time credit in points for the same sort of tested ability tends to lessen the prediction value of the combined score. It decreases the value which the supplementary abilities would have in the total prediction. Rosenow (30) discusses this question at length in weighting the Kitson tests at Chicago University. Two features of the data which Hardwick presents are important, although they are not commented on by her. Five of the tests show a difference of .50 or more in the different age groups in the closeness of their correlation with the total score. Moreover, between the ranking of the tests on the basis of high correlation for the 6-year-olds and for the 12- and 13-year-old group there is a minus correlation. This suggests that a weighting for younger ages probably would not hold for older ages. High correlations with general intelligence often indicate a wider scatter of the test scores; but not theoretically, as she assumes, a nearer approach to a normal distribution of scores.

Method.—Rugg's chapter on correlation (31) presents in 75 pages for the non-mathematical reader the best simple treatment of the measurement of relationship. Schoolmen will especially appreciate his success in explaining the reasons for the principal devices and their significance without expressing the explanation in technical language. The methods of computing the various coefficients are also presented simply and fully. Unfortunately, space forbade his including partial correlation which illuminates the whole subject of analysis of causes. It will probably be the most important feature of the work in correlation during the next decade. An admirable illustration is given of the non-linear relation of cost-per-student-instructed and number of pupils instructed by one teacher. Diagram 43 would be simpler and more in accord with Galton's presentation in his *Natural Inheritance* if the line which measures the proportion of perfect correlation extended from the diagonal of perfect correlation across the regression line perpendicular to the mean X axis.

Otis (26) has shown how to find the relative weighting of two variables for predicting a third without calculating any of the partial correlation coefficients. The method saves about half the calculations of the usual partial regression equation and is useful when the errors of the prediction are not required. He also provides the corresponding formula for more than three variables, which reduces the number of partial coefficients to be calculated. Rosenow (30)

contributes a scheme for the mechanical technique of calculating partial coefficients which may reduce the labor of calculation as much as half, when only the relation to one dependent variable is required as is common in psychological work.

The normal probability table does not apply accurately to the probable errors of coefficients of correlation calculated from small samples. An elaborate co-operative study (35) now makes available further tables for estimating the skewed distribution surfaces of the correlations for small samples. These surfaces are all forms from a rectangle to J and U forms, but in general it may be said that the distributions of the correlations in small samples tend to have modes above the true correlations and means below the true correlations. For example, with a true correlation of .50 in the population sampled, the modal correlation expected, when calculated from samples of 10 cases, would be .63 and the expected mean of the correlations in such samples, .48. Pearson (27) provides the tables for determining the error of the correlation coefficient calculated by his biserial η , when one variable is given in alternative categories and the other by multiple categories. Lee (18) completes Everitt's work by providing the supplementary tables for calculating correlations above plus or minus .80 by the tetrachoric method in certain cases not previously covered. We may expect more and more use of this method with social material in which the data can often be divided only into two groups.

Analysis of Abilities.—The use of correlation for unraveling the factors involved in various abilities continues to be fruitful. In this connection, the study of special abilities is suggestive. Seashore and Mount (34) give an elaborate analysis of the relations of tested musical talent to musical training, environment, the enjoyment of music, and musical accomplishment. The low correlations indicate that training is commonly given with little regard to talent. Stevenson (36) finds that intercorrelations of discrimination of sounds, lines and brightnesses, when accurately measured, bring out close relationships (.90 plus) in sense discrimination. With pressure, accurate tests were not possible and the correlation was less. In a penetrating study Downey (9) shows the possibility of the degree of unidextrality being indicative of specialized capacity which may be of value in some forms of spatial orientation. It is related to mirror-reading and, perhaps, to visualization. Gordon (11) does not find that unidextrality goes with either strength of grip or maturity. The correlations between abilities in the four

fundamental operations in arithmetic are from .46 to .75. McQueen (21) by a very complete statistical treatment, including partial correlation to eliminate the effect of certain factors, finds that there is no evidence of a general power of distribution of the attention. Bickersteth (2) also gives negative evidence for common factors. The small correlations between tests for divided and sustained attention give no support for the idea of one capacity of attention. Rosenow (30) illustrates with Kitson's test data on Chicago University students what may be done by correlation analysis of mental functions. Bennet (1) found auditory and visual presentation with disconnected materials had more in common than with connected materials, and that ability in mediate retention has more in common with different materials than ability for immediate recall. Gordon (12) shows an independence of memory for nonsense syllables and rapid learning of music. Dearborn and Brewer (7) give the relations of a class experiment in learning a code to tests of memory. Buckingham (5) finds correlations of about .4 between tests for ability to think about and ability to remember facts in history.

In Business Psychology.—Hollingworth and Poffenberger (15) give the correlations between tests and success in 11 occupations; Burt (6) gives them for the Münsterberg vocational tests; Rogers (29) for stenographers and typewriters. The latter found that the empirical selection of tests would give "a far more reliable criterion for vocational guidance in the field of stenography than has ever been attained by any other method." Moody (24) gives the correlation between salaries of teachers and their academic grades. Marks in the theory and practice of teaching are no better than average scholarship. Inspection of the tables shows closer relations with low marks and low salaries. The highest coefficient found with salaries was .52. A method of estimating various personal traits for the use of the employment bureau at the Carnegie Institute of Technology is evaluated by Miner (22). Estimates of leadership are least indicated by grades. Gould (13) shows that a test for memory of names and faces agrees .75 with the empirical test of this ability among 10 psychologists when trying to remember a group of 17 students after brief introductions.

Mental Development.—Correlation was used extensively by Bickersteth (2) in his study of the variations of tested abilities with maturity. The same 12 tests were given to children of different ages. He thinks that the tests reached capacity because they

showed small correlations with life-age, when children were in the same grades and were from 9 to 11 in age. There is small relation between mental and motor ability and it decreases with age, while the interrelations of the tests of higher mental traits increase with age. In an important comparison between normal and feeble-minded children of the same mental ages, Woodrow (39) concludes that there is an absence of correlation between capacity to learn and capacity to grow mentally. The correlations for absolute and for relative gain with practice in a form sorting test for the normal and for the feeble-minded group were negative with initial ability in the test. Those for percentage of improvement were practically three times their P. E.'s. This is only apparently contrary to the usual results in which those who are best in a test improve most with subsequent practice since here differences in mental age were almost eliminated. Following Woodrow's argument we might suppose that greater mental ability, even greater mental maturity; but not greater capacity for mental growth (capacity to increase mental age) may go with rapid learning. "The connection between language differences and intelligence differences becomes more intimate with increasing years," so Brigham found (4). Terman *et al.* (38) found that the correlation between I Q's obtained with the Stanford Scale and teacher's estimates of intelligence increased from .48 to .71 when the effects of extraneous factors upon the estimates were eliminated. Doll (8) shows that the results with his two brief Binet scales correlate over .90 with the complete mental age ratings for either normal or feeble-minded children. The ratio of actual to expected school grade, Pedagogical Quotient, correlates .82 with the I Q's of school children. Race (28) gives the correlations of I Q's with various tests. The opposites test is closest related and Curtis' Arithmetic test only .09.

School Records.—Besides the studies of Ruml (32) and Rosenow (30), there are numerous correlations with school work. The closest relation for a series of tests and school grades was at the Whitewater Normal where James (16) found that the average record from his series of five tests correlated .55 to .63 with different school subjects. Bright students not studying and dull students studying hard reduce the correlations. Gordon (11) found a correlation of .77 between a mechanical syllogism test and college grades. King and M'Crory (17) give correlations for 276 women and 268 men in the freshman class of the College of Liberal Arts at Iowa University. The intercorrelations of the tests are also given, but

the authors do not combine them in a multiple coefficient. The differences between the correlations with university grades for men and for women, especially with the opposites test, .84 and .45 respectively, suggest that the male group is more varied in ability. It is important to remember that the size of the coefficient increases decidedly with increase in the extent of scatter of the group. This may account also in part for the low correlations of tests with class averages found by Sunne (37) and also for those by Garrison (10) who used scores in the Point Scale. The tests often do not sufficiently discriminate those of best ability. Minnick (23) finds that four tests for different types of ability in geometry show widely different correlations in different schools. How much this result is influenced by differences in the native variability of pupils in the different geometry classes and how much is due to difference in methods of teaching is an important educational problem. Coefficients of .55 or over were found for each test in some school. Myers (25) found that in delayed recall students remembered best the names of the military men most frequently mentioned in histories, but that this was not true for names of prominent civilians. Breed (3) found that Starch's method of testing the comprehension of reading does not measure the same factors as Thorndike's. McCall and Ruger (20) demonstrate that a study of the correlation of psychological and educational measurements was closely duplicated by a similar study a year later. Lincoln (19) finds that high school marks correlate better with college marks than do entrance examinations.

REFERENCES

1. BENNETT, F. The Correlation between Different Memories. *J. of Exp. Psychol.*, 1916, 1, 404-419.
2. BICKERSTETH, M. E. The Application of Mental Tests to Children of Various Ages. *Brit. J. of Psychol.*, 1917, 9, 23-73.
3. BREED, F. S. A Comparison of Two Methods of Measuring Comprehension in Reading. *School & Soc.*, 7, 266-270.
4. BRIGHAM, C. C. Two Studies in Mental Tests. *Psychol. Monog.*, 1917, 24, No. 1. Pp. 254.
5. BUCKINGHAM, B. R. Correlation between Ability to Think and Ability to Remember, with Special Reference to United States History. *School & Soc.*, 1917, 5, 443-449.
6. BURT, H. E. Professor Munsterberg's Vocational Tests. *J. of Appl. Psychol.*, 1917, 1, 201-213.
7. DEARBORN, W. F., & BREWER, J. M. Methods and Results of a Class Experiment in Learning. *J. of Educ. Psychol.*, 1918, 9, 63-82.
8. DOLL, E. A. A Brief Binet-Simon Scale. *Psychol. Clinic*, 1918, 11, 197-261.
9. DOWNEY, J. E., & PAYSON, E. B. Unidextrality and Mirror-Reading. *J. of Exp. Psychol.*, 1917, 11, 393-415.

10. GARRISON, S. C. Yerkes Point Scale for Measuring Mental Ability as Applied to Normal Adults. *Psychol. Bull.*, 1917, 14, 152-153.
11. GORDON, K. *Educational Psychology*. New York: Holt, 1917. Pp. vi + 294.
12. GORDON, K. Some Tests on the Memorizing of Musical Themes. *J. of Exp. Psychol.*, 1917, 2, 93-99.
13. GOULD, R. L. A Test for Memory of Names and Faces. *J. of Appl. Psychol.*, 1917, 1, 321-324.
14. HARDWICK, R. S. Weighting of Point Scale Tests. *J. of Educ. Psychol.*, 1917, 8, 416-424.
15. HOLLINGWORTH, H. L., & POFFENBERGER, A. T. *Applied Psychology*. New York: Appleton 1917. Pp. xiii + 337.
16. JAMES, B. B. Correlations of Mental Tests and Scholarship. *School & Soc.* 1918, 7, 238-239.
17. KING, I., & M'CRODY, J. Freshmen Tests at the State University of Iowa. *J. of Educ. Psychol.*, 1918, 9, 32-46.
18. LEE, A. Further Supplementary Tables for Determining High Correlations from Tetrachoric Groupings. *Biometrika*, 1917, 11, 284-291.
19. LINCOLN, E. A. The Relative Standing of Pupils in High School, in Early College, and on College Entrance Examinations. *School & Soc.*, 1917, 5, 417-420.
20. MCCALL, W. A., & RUGER, G. J. Reliability of Ph.D. Research Dissertations in Educational Psychology. *School & Soc.*, 1918, 7, 441-449.
21. McQUEEN, E. N. Distribution of Attention. *Bru. J. of Psychol., Monog. Supplements*, 1917, 5. Pp. vii + 142.
22. MINER, J. B. The Evaluation of a Method for Finely Graduated Estimates of Abilities. *J. of Appl. Psychol.*, 1917, 1, 123-133.
23. MINNICK, J. H. Certain Abilities Fundamental to the Study of Geometry. *J. of Educ. Psychol.*, 1918, 9, 83-90.
24. MOOI Y, F. L. Correlation of the Professional Training with the Teaching Success of Normal School Graduates. *School Rev.*, 1918, 3, 180-198.
25. MYERS, G. C. Delayed Recall in History. *J. of Educ. Psychol.*, 1917, 8, 275-283.
26. OTIS, A. S. The Derivation of Simpler Forms of Regression Equations. *J. of Educ. Psychol.*, 1917, 9, 619-621.
27. PEARSON, K. On the Probable Error of Biserial Eta. *Biometrika*, 1917, 11, 292-302.
28. RACE, H. A Study of a Class of Children of Superior Intelligence. *J. of Educ. Psychol.*, 1918, 9, 91-98.
29. ROGERS, H. W. Psychological Tests for Stenographers and Typewriters. *J. of Appl. Psychol.*, 1917, 1, 268-274.
30. ROSENOW, C. The Analysis of Mental Functions. *Psychol. Monog.*, 1917, 24. Pp. 43.
31. RUGG, H. O. *Statistical Methods Applied to Education. A Textbook for Students of Education in the Quantitative Study of School Problems*. Boston: Houghton Mifflin, 1917. Pp. ix + 410.
32. RUMI, B. Coefficients of Diagnostic Value. *J. of Phil., Psychol. & Sci. Meth.*, 1917, 14, 633-637.
33. RUMI, B. Reliability of Mental Tests in the Division of an Academic Group. *Psychol. Monog.*, 1917, 24. No. 4.
34. SEASHORE, C. E., & MOUNT, G. H. Correlation of Factors in Musical Talent and Training. *Psychol. Monog.*, 1918, 25, 47-92.

35. SOPER, H. E.; YOUNG, A. W.; CAVE, B. M.; LEE, A.; & PEARSON, K. On the Distribution of the Correlation Coefficient in Small Samples. Appendix II to Papers of "Student" and R. A. Fisher. *Biometrika*, 1917, 11, 328-413.
36. STEVENSON, J. A. Correlation between Different Forms of Sensory Discrimination. *J. of Appl. Psychol.*, 1918, 2, 26-42.
37. SUNNE, D. The Relation of Class Standing to College Tests. *J. of Educ. Psychol.*, 1917, 8, 193-211.
38. Terman, L. M., Lyman, G., Ordaahl, G., & Ordaahl, L. E., Galbreath, Neva, and Talbert, Wilford. *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence*. *Educ. Psychol. Monog.*, 1917, No. 18. Pp. 179.
39. Woodrow, H. Practice and Transference in Normal and Feeble-Minded Children. *J. of Educ. Psychol.*, 1917, 8, 85-96, 151-165.