THE SHRINKAGE OF THE BROWN-SPEARMAN PROPHECY FORMULA

By Robert J. Wherry

At the recent meeting of the Conference on Individual Psychological Differences held in Washington, Dr. Clark Hull of Yale University called attention to the fact that the much used Brown-Spearman formula involves, or leads to, if used without regard to certain limitations, a certain over optimism.[1] In other words, if only this formula is taken into account, one would assume that the mere increasing in length of a test would automatically and, with continued increases in length, *indefinitely* continue to increase its reliability or validity.

On the other hand, we know that the greater the number of test units the greater the shrinkage between the predicted and actually obtained value. At least we know this to be true when the value in question is a multiple correlation coefficient and the test units are independent variables. Hull raised the question as to whether or not the same fact might be true of the figures predicted by the Brown-Spearman formula. It is the purpose of this article to show that this shrinkage does occur, and that the Wherry-Smith shrinkage formula[2] satisfactorily predicts this shrinkage.

A quick review of the nature of the two formulae (the Brown-Spearman and the Wherry-Smith formulae) will at once show the importance of the discussion. The Brown-Spearman formula, as applied to the predicting of reliability, reads as follows,

$$R = \frac{M\, r_{11}}{1 + (M - 1)\, r_{11}}, \tag{1}$$

where R = the predicted reliability,
r_{11} = the discovered reliability,
and M = the number of times the test is lengthened. Thus the test provides that the predicted reliability (R) will increase with each increase in M, but it is to be noted that the increase in R decreases with each increase in M as the value of R approaches its limit of plus one.

On the other hand the Wherry-Smith formula, which reads,

$$\bar{R}^2 = \frac{(N - 1)R^2 - (M - 1)}{N - M}, \tag{2}$$

where $\bar{R}$ = the predicted value of the correlation,
R = the discovered correlation,
M = the number of independent variables

and N = the statistical population (the number of cases), provides that, for each increase in M, the shrinkage in $\bar{R}$ as compared with R increases. Thus, if

TABLE I

Correlations Observed and Theoretical (Based upon Observed Means)
(N = 37 throughout)

M	Observed average	Correlation predicted		Error	
		Br.-Sp.	Wherry	Br.-Sp.	Wherry
	(Trait 1)				
1	.290				
5	.728	.671	.618	−.057	−.110
10	.717	.803	.726	.086	.009
15	.754	.860	.758	.106	.004
20	.805	.891	.825	.087	.020
30	.936	.925	.509	−.011	−.427
	(Trait 5)				
1	.419				
5	.736	.783	.751	.047	.015
10	.845	.878	.834	.033	−.011
15	.887	.915	.856	.028	−.031
20	.877	.935	.856	.058	−.021
30	.876	.956	.745	.080	−.131
	(Trait 10)				
1	.354				
5	.479	.733	.692	.254	.213
10	.717	.846	.788	.129	.071
15	.852	.892	.816	.040	−.036
20	.636	.915	.822	.279	.186
30	.805	.943	.655	.138	−.150
	(All Traits)				
1	.320				
20	.898	.904	822	.006	−.076
30	.872	.933	.576	.061	−.296

we assume that the M's in the two formulae are analogous, i.e., if we assume the Wherry-Smith formula to be applicable to the Brown-Spearman formula, we see that as M increases the Brown-Spearman formula adds a decreasing increment while the Wherry-Smith formula provides that an increasing decrement be subtracted, thus eventually we arrive at a point where by further increasing the length of the test we will decrease rather than increase the size of the reliability coefficient.

If our hypothesis be true, we must, then, in order to predict the correct value of $\bar{R}$, substitute the value of equation (1) in equation (2). Doing this we have

$$\bar{R}^2 = \frac{(N-1)M^2r_{11}^2 - (M-1)^3r_{11}^2 - 2(M-1)^2r_{11} - (M-1)}{(N-M)[1 + 2(M-1)r_{11} + (M-1)^2r_{11}^2]}, \qquad (3)$$

which would then be the form in which the Brown-Spearman formula should be used in predicting reliability corrected for chance error by the Wherry-Smith

TABLE II

Error in Predicting Reliability (Based upon Observed Means)

Error	Brown-Spearman	Wherry
over .210	2	1
.151– .210		1
.091– .150	3	
.031– .090	8	1
−.029– .030	3	6
−.089– −.030	1	3
−.149– −.090		3
−.209– −.150		
below −.209		2

TABLE III

Rietz Criteria of Normality Applied to Results from Means

Criterion	Normal	Brown-Spearman	Wherry
u_1	0	.074	−.032
β_1	0	.561	−.283
β_2	3	2.008	3.180

formula. The same result can of course be secured by applying the formulae consecutively.

In order to test the formula (3), the writer has applied it to some empirical data. A recent article by H. H. Remmers of Purdue University furnishes the needed data. Remmers study dealt with the increase in reliability due to increase in the number of judgments of certain traits of college professors.[3] His results, together with the results of applying formula (3) to the data are shown in Table I.

An inspection of Table I shows at once that while the Brown-Spearman formula gives results which are consistently too large (15 out of 17 times) the Wherry-Smith formula gives results which are more nearly equally distributed

TABLE IV

Correlations Observed and Theoretical (Based upon Observed Medians)

(N = 37 throughout)

M	Observed medians	Correlation predicted		Error	
		Br.-Sp.	Wherry	Br.-Sp.	Wherry
	(Trait 1)				
1	.344				
5	.752	.724	.682	−.028	−.070
10	.663	.840	.779	.177	.116
15	.702	.887	.807	.185	.105
20	.805	.913	.805	.108	.000
30	.936	.940	.635	.004	−.301
	(Trait 5)				
1	.450				
5	.760	.804	.776	.040	.016
10	.856	.891	.852	.035	−.004
15	.931	.925	.873	−.006	−.058
20	.877	.942	.874	.065	−.003
30	.876	.961	.778	.085	−.098
	(Trait 10)				
1	.363				
5	.433	.740	.701	.307	.268
10	.754	.851	.795	.097	.041
15	.872	.895	.822	.023	−.050
20	.898	.919	.820	.021	−.078
30	.872	.945	.669	.073	−.203
	(All Traits)				
1	.503				
20	.898	.953	.879	.055	−.019
30	.872	.968	.829	.986	−.043

between positive and negative errors (7 to 10), tending to slightly underestimate. The actual distribution of errors can be more easily seen by an inspection of Table II.

Now, if our formula were perfectly correct, we should expect that the errors incurred by its use would be normally distributed about a mean error of zero. The Rietz criteria for normality of distribution were applied to these errors with results as shown in Table III.[4] It can be readily seen that the Wherry correction formula gave much better results than did the uncorrected Brown-Spearman formula when measured by the Rietz criteria.

All of the results in the first three tables are based upon the means of the results obtained by Remmers, since this was the method used in his paper. However, when the number of cases is small, as they were in this study, it is

TABLE V

Error in Predicting Reliability (Based upon Observed Medians)

Error	Brown-Spearman	Wherry
over .210	1	1
.151– .210	2	
.091– .150	3	2
.031– .090	5	1
−.029– .030	6	5
−.089– −.030		5
−.149– −.090		1
−.209– −.150		1
below −.209		1

TABLE VI

Rietz Criteria of Normality Applied to Results from Medians

Criterion	Normal	Brown-Spearman	Wherry
u_1	0	.074	−.018
β_1	0	.497	−.081
β_2	3	1.599	2.284

sometimes preferable to use the median rather than the mean as a basis of calculation, since the median is less affected by extreme cases. The writer has therefore recalculated the problem on the basis of the medians discovered by Remmers, and the results are given in Tables IV, V, and VI. The results were found to differ but little from those based upon the means of the distributions.

If we now assume that the formula (3) has been empirically established and justified, we must still answer a very practical question, namely, "How long shall we make our tests in order to achieve the greatest reliability?" To answer this question we must find the point at which $\bar{R}$ becomes a maximum, with respect to changes in M, assuming r_{11} and N to be constant terms. To find this point we must find the derivative of equation (3) with respect to M and set the numerator equal to zero, thus, if we write Formula (3) in a slightly more usable form, we have,

$$\bar{R}^2 = \frac{(N-1)M^2 r_{11}^2}{(N-M)(1+2[M-1]r_{11}+[M-1]^2 r_{11}^2)} - \frac{M-1}{N-M}, \tag{3a}$$

whence

$$\frac{d\bar{R}^2}{dM} = \frac{(1+[M-1]r_{11})\{4r_{11}^2 M^2 - (2Nr_{11}^2 + 3r_{11}[1-r_{11}])M + (1-r_{11})^2\}}{(N-M)^2(1-2[M-1]r_{11}+[M-1]^2 r_{11}^2)^2} \tag{4}$$

which causes $\bar{R}$ to reach a maximum or minimum when the numerator is placed

TABLE VII

Showing the value of M which will give a maximum value for $\bar{R}$

(According to the Brown-Spearman-Wherry-Smith formula)

N	r_{11}								
	.10	.20	.30	.40	.50	.60	.70	.80	.90
10	Imag.	Imag.	3	4	4	4	5	5	5
20	Imag.	6	8	9	9	10	10	10	10
30	Imag.	12	13	14	14	14	15	15	15
40	11	17	18	19	19	19	20	20	20
50	17	22	23	24	24	24	25	25	25
60	22	27	28	29	29	29	30	30	30
70	27	32	33	34	34	34	35	35	35
80	32	37	38	39	39	39	40	40	40
90	38	42	43	44	44	44	45	45	45
100	43	47	48	49	49	49	50	50	50

equal to zero. Thus, placing the numerator equal to zero and factoring this equation, we find its roots to be

$$M = \frac{-(1-r_{11})}{r_{11}} \tag{5a}$$

$$M = \frac{2Nr_{11} - 3(1-r_{11}) - \sqrt{4N^2 r_{11}^2 - 12Nr_{11}(1-r_{11}) - 7(1-r_{11})^2}}{8r_{11}} \tag{5b}$$

or

$$M = \frac{2Nr_{11} - 3(1-r_{11}) + \sqrt{4N^2 r_{11}^2 - 12Nr_{11}(1-r_{11}) - 7(1-r_{11})^2}}{8r_{11}} \tag{5c}$$

and by substituting actual values of N and r_{11} in the equations, we find that equation (5c) is the root we are seeking (i.e.) the value of M for which $\bar{R}$ becomes a maximum.

It can also be readily seen that the value under the radical approximates a perfect square (lacking 16 units of being that figure) of the quantity outside of the radical, thus approximating this value for large values of N. Thus, when N is large (exceeds 100) we may secure satisfactory approximations to M if we rewrite equation (5c) in the form below

$$M_{(\text{Approximately})} = \frac{N}{2} - \frac{3(1 - r_{11})}{4r_{11}}. \tag{5d}$$

Table VII shows the results of equation (5c) for values between $N = 10$ and $N = 100$ (by increments of 10) for values of r_{11} from .10 to .90 (by increments of .10). The use of the formula does not yield integers, and so the results in the table are recorded to the nearest whole number rather than exactly as given by the formula.

If, in order to test the validity of formula (5c), we apply it to the values in Tables I and IV, we find fairly close agreement. The formula in each case predicts a maximum value for $\bar{R}$ when M lies between 15 and 20, and in the actually lengthened tests R is found to be a maximum when M is 30, 15, 15, 20, 30, 15, 20, and 20, thus being in agreement six times out of eight.

Conclusions

1. The Brown-Spearman formula appears to give results which contain both constant and chance errors.

2. These results can be practically eliminated by applying the Wherry-Smith correction formula to the results obtained by the Brown-Spearman formula.

3. We may find the value of M which will give the greatest value of $\bar{R}$ by substitution in equation (5c) above, and then by substitution of this value in equation (3), find the most probable value of $\bar{R}$ at its maximum point.

4. For large values of N we may secure satisfactory approximations to M by means of the simpler formula (5d).

BIBLIOGRAPHY

1. Hull, Clark: "Memorandum Concerning Factors Influencing the Prediction of Performance," Appendix F, Conference of Individual Psychological Differences, National Research Council, Washington, D. C., 1930.
2. Wherry, Robert J.: "A New Formula for Predicting the Shrinkage of the Multiple Correlation Coefficient," Annals of Mathematical Statistics, November, 1931.
3. Remmers, H. H.: "The Equivalence of Judgments to Test Items in the Sense of the Spearman-Brown Formula," Journal of Educational Psychology, January, 1931.
4. Rietz, H. L.: Mathematical Statistics, Carus Mathematical Monograph, No. 3, Chicago, 1927, pp. 58–59.

Cumberland University
Lebanon, Tennessee