Proceedings of the Iowa Academy of Science

Volume 45 | Annual Issue Article 68

1938

Concerning the Establishment of a Criterion and Certain Incidental Points Relations to Validity Coefficients

A. R. Lauer
Iowa State College

CONCERNING THE ESTABLISHMENT OF A CRITERION AND CERAIN INCIDENTAL POINTS RELATING TO VALIDITY COEFFICIENTS

A. R. LAUER

There is a persistent demand for validity coefficients of some sort in the field of applied psychology. The simplest criterion is often chosen. The Binet-Simon test has been validated almost entirely against an academic criterion. This may or may not be justified as there are many who believe that intelligence is more than mere scholastic aptitude. With more recent explorations, by factor analysis and other techniques, this belief has considerable experimental support. We have thus sought for a criterion and too often have been content with that which gives the highest validity coefficients.

When the more complex phenomena are considered, however, these coefficients drop to astonishingly low points. It has been the general practice to discard all test devices which show low validity coefficients. Sometimes this is unjustified as it does not follow that there is no relationship because a correlation coefficient is zero. Eta coefficients and partial correlation may often reveal obscured relationships. But, assuming there are no statistical "stumbling-blocks" to trip us up; suppose a correlation with a criterion is low. Does this mean there is no relationship? It may and it may not, depending upon the complexity of the criterion and other considerations. It is quite well-known that most aptitude tests have low validity. In studies of automobile drivers we have obtained no high coefficients of correlation with accidents. It is on this point that the present paper hinges. Why do we not get higher validity from reliable tests?

DEFECTS IN CRITERIA

Further investigation has revealed a large number of defects in fundamental assumptions usually made concerning criteria, as well as in the criteria proper. Some of these are:

1. It is often assumed that any one test of some specific ability will corrrelate highly with success in the function involving that ability. This is not a valid assumption for several reasons.
   a. Numerous associated factors tend to dilute the results obtained, which spuriously lower the correlation.

b. The human organism may compensate for weaknesses present and thus change the effect of a deficiency, thereby still lowering the correlation with the criterion.

2. It is often assumed that one category will suffice as a criterion. For example, in automobile driving the number of accidents may be set up as a criterion. This is a fallacious premise for the following reasons:
   a. The accident record may have been of long standing and factors which were potent at the time of the accident are no longer operative.
   b. Accident statistics are woefully inadequate. There is no such thing as determining proximate causes in vehicular accidents with any degree of accuracy.
   c. Several contributing factors are usually operative at the same time. It has been shown that at least three distinct causes may be ascribed to every accident.
   d. Often no provision is made to correct for the mileage driven, number of automobiles used, conditions of driving and years of driving experience.
   e. Even when some of these corrections are made they tend to neutralize each other. For example, there is a negative correlation between miles driven each year and susceptibility to accidents as measured by the actual number of accidents the driver has, per unit of time. The exposure record, in general, will correlate positively with accidents; i.e., the greater the risk the more numerous the accidents.
   f. There are occasional breaks in equipment and conditions arise which no driver could avert. Consequently there will be accidents — a certain few — which are unavoidable.
3. It is commonly believed by practical statisticians that a correlation below .25 is not worth while since the amount of perfect prediction is somewhat less than four per cent. Such interpretation is misleading for the following reasons:
   a. The percentage of factors measured is much higher, being about 20.5 according to Table I.
   b. The calculation of percentage of perfect prediction overlooks the "near-hits" in favor of "bulls-eyes." In other words a "miss" is assumed to be a "random shot" instead of "a runner-up."
   c. There is something wrong with the methods used to determine the percentage of perfect "hits." Theory does not check with empirical facts demonstrated in the next section. Why is this true?

AN EMPIRICAL STUDY

This question was subjected to an experimental investigation. Table I shows theoretical data from a standard source. By glancing down the second column it is obvious that low validity coefficients are discouraging to those who wish exact results in application.

Table I

(Excerpt from published report based on Minor's Tables)

Coefficient of Correlation	Percentage of Perfect Prediction	Percentage of Determining Factors Measured
1.00	100.0	100.0
.95	68.8	75.3
.90	56.4	67.4
.85	47.3	62.7
.80	40.0	57.1
.75	33.9	53.1
.70	28.6	49.5
.65	24.0	46.1
.60	20.5	42.9
.55	16.6	39.7
.50	13.4	36.6
.45	10.7	33.5
.40	8.3	30.4
.35	5.6	27.2
.30	4.6	23.9
.25	3.2	20.5
.20	2.0	16.9
.18	1.1	13.2
.10	0.5	9.1
.05	0.1	4.8
.00	0.0	0.0

To determine the validity of this table, telephone numbers were drawn from the directory, odds versus evens, and correlated. The results are obviously low as one might expect since no relationship exists. The correlation was actually —.0136.

Next a known number of perfect relationships were superimposed upon the principal diagonal of the plot. It was found that the number of perfect predictions closely approximates the coefficient of correlation. Some of the earlier studies were made with the population evenly distributed over the quadrant.

While these data are relatively incomplete, the consistency of results points to a grave flaw in the theoretical tables usually given on this point. It suggests a complete review of published accounts of the significance of a correlation coefficient, and perhaps the need for empirical verification of theoretical considerations. While the perfectly related items were laid out on the diagonal, it indicates the possibilities of correlation with a criterion. Compare, for example, the test of glare as correlated with accidents as a whole. It is estimated that 7 per cent of all accidents which happen at night are due to glare. About 60 per cent of all accidents occur at night. Then 60 per cent of 7 would give 4.2 per cent of perfect prediction due to this factor, assuming a correlation of unity between accidents said to be caused by glare and susceptibility to glare. The latter assumption, of course, is absurd. Thus reading

Table II
(Effects of Known Amounts of Superimposed Data)

Plot N 100	N Total	Original Correlation with 100 Cases	Percentage added on Diagonal Above Chance	Final Correlation
Even distribution with 10 on diagonal	110	00	9.2	.0909
Even distribution with 20 on diagonal	120	00	16.6	.1666
Even distribution with 30 on diagonal	130	00	23.0	.2300
Even distribution with 40 on diagonal	140	00	28.6	.2860
Even distribution with 50 on diagonal	150	00	33.5	.3330
Even distribution with 60 on diagonal	160	00	37.5	.3750
Even distribution with 70 on diagonal	170	00	41.1	.4110
Even distribution with 80 on diagonal	180	00	44.4	.4440
Even distribution with 90 on diagonal	190	00	47.3	.4730
N 110				
Telephone numbers with 9 on diagonal	119	—.0136	8.2	.0285
Telephone numbers with 18 on diagonal	128	—.0136	16.3	.1330
Telephone numbers with 27 on diagonal	137	—.0136	19.7	.1900
Telephone numbers with 36 added on diagonal	146	—.0136	24.7	.2420

back to the tables cited it would be impossible to obtain a correlation higher than +.042 between glare susceptibility and all accidents lumped together. According to most statisticians such a validity coefficient would be practically worthless.

SUMMARY AND CONCLUSIONS

From both theoretical and empirical arguments it is concluded that validity coefficients of specific tests used against a complex criterion cannot be high. Satisfactory validity coefficients may be obtained only by complete analysis of the criterion into functions which are quite like the tests themselves. If this cannot be done the search for high validity will be futile. In guidance work *it will* thus be necessary to accurately job-analyze the requirements of the function before any satisfactory prognostic index can be attained. The interrelation of variables must also be obtained in order to show up obscured relationships.

DEPARTMENT OF PSYCHOLOGY,
IOWA STATE COLLEGE,
AMES, IOWA.