Multiple regression and multiple correlation

After we have reduced our battery, either by eliminating tests or combining tests, to the point where linear restraints are no longer present, we may determine the "best" weights by the multiple regression procedure. In doing so, we are finding the weights which are "best" for the given sample. The least squares procedure gives the best possible fit to the sample statistics, including their sampling errors. The multiple correlation in the sample is as high as it can be for that given set of individuals. If we draw a second sample of persons from the same population, give them the same tests, and secure similar criterion scores, we shall have to use the regression coefficients from this second sample to obtain a multiple correlation coefficient which is equally likely to be above or below the one obtained from the first sample. If we apply the regression weights from the first sample to the scores from the second, the resulting aggregate correlation in the second sample will necessarily be lower. But this is exactly what we do when we use a test battery for any practical purpose. We determine the regression weights on the basis of statistics from an experimental sample. Then we use them in making predictions concerning the criterion behavior of *other* persons from the same population.

The predictive power or validity of a battery must, therefore, be determined by giving it to a second sample from the same population as the sample used in determining the weights. Using the weights from the first sample, we determine the aggregate correlation in the second sample. This will almost always be substantially lower than the multiple correlation in the first sample, and it will necessarily be lower than the multiple correlation in the second sample. This aggregate correlation, and not the multiple correlation in either sample, is the index of prediction for the battery. When divided by the index of reliability of the criterion from the second sample, it becomes the index of validity of the battery. This procedure is termed "cross-validation." It is the only procedure which provides an uninflated estimate of the predictive power or validity of a weighted battery.

Item analysis, test selection, and validity

These same considerations apply with even greater force when we are dealing with single items instead of tests. Whether the procedure be item selection or item weighting or both, we are dealing partly with real relationships and partly with chance. The selected items are the ones which are best in the given sample. The item weights are also the ones which are best in that sample. The validity of the reduced and/or weighted test