EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT
1985, 45

COMPARISON OF FOUR METHODS FOR WEIGHTING MULTIPLE PREDICTORS

MICHAEL G. AAMODT
Radford University

WILSON W. KIMBROUGH
University of Arkansas

An important issue in personnel psychology concerns the methods that are used to combine multiple criteria and/or multiple predictors into a single composite. Fralicx and Raju (1982) looked at five methods of combining multiple criteria and concluded that with the exception of canonical weights, the results obtained from four of the methods were almost identical. The present study differed from Fralicx and Raju (1982) in that it used predictors rather than criteria, was composed of a different type of sample, used different types of weighting methods, a smaller number of dimensions, and included a cross-validation procedure. In spite of these differences the conclusion drawn in both studies is similar. That is, most weighting methods are highly related. Furthermore, the present study demonstrated that no weighting method was superior to another in terms of protection from validity shrinkage.

ONE of the more common procedures in making personnel selection decisions is to take a number of predictors and combine them into a composite score that can be used to evaluate overall differences in job candidates. To do this, it is necessary to decide upon the most appropriate method to use in assigning weights to the separate predictors forming the composite. While there are many possible methods (regression weights, unit weights, etc.) that can be

The authors would like to thank Robert J. Keller, Krista J. Crawford, and the University of Arkansas Housing Department for their help in this study.

Requests for reprints should be addressed to Michael G. Aamodt, Department of Psychology, Radford University, Radford, VA 24142.

used to weight predictors, there is little information available as to which is the best method to use. While use of regression weights is perhaps the most popular method of weighting predictors (McCormick and Ilgen, 1980), empirical arguments have been made (Schmidt, 1971; Einhorn and Hogarth, 1975) that use of unit weights is equal to, and in terms of validity shrinkage, better than regression weights. Furthermore, Lawshe (1969) has shown that the differential weighting of predictors leads to validity coefficients no higher than the adding of raw scores.

Recently, support for the notion that most methods of weighting scores, in this case criterian scores, are equal was found by Fralicx and Raju (1982) when they compared five methods for combining multiple criteria into a single composite. Fralicx and Raju (1982) correlated the weights derived by management importance ratings, equal weighting (multiplying a criterion variable by the reciprocal of it's standard deviation), unit weighting, factor weights, and canonical weights. The results indicated that the first four methods correlated near unity with one another while the canonical weights correlated near zero with the other weights.

While the findings of the Fralicx and Raju (1982) study are certainly important and suggest that it may not really matter as to which method is used to weight criteria, the study is limited in that it only looked at the intercorrelations between various methods. It is possible that even though two methods are highly correlated, one method could still be superior to the other method. In fact, this is what Schmidt (1971) found with the use of regression coefficients as a method of weighting multiple predictors. That is, unit weights were less susceptible to validity shrinkage during cross-validation than were regression weights.

It was the purpose of the present study to examine the relationship among four methods for combining predictors into a single composite and furthermore, to determine if one method is superior to another in terms of resistance to validity shrinkage during cross validation. To do this, two groups of University Housing Resident Assistants were used. The first group participated in a concurrent validity study during Fall, 1981 while the second group participated in a cross-validation study during Spring, 1982. Both groups were administered the Trait Evaluation Index (Nelson, 1966) and the scores from two of the scales were weighted in four different ways and then summed to form a composite score. It was expected that, consistent with the results of Fralicx and Raju (1982), the composite scores from the four different methods would all be highly correlated and that similar to Schmidt (1971), unit weights would lead to less validity shrinkage than would regression weights.

Method

Subjects

The subjects in the concurrent validity study were 62 Housing Resident Assistants at the University of Arkansas while the subjects in the cross-validation sttdy were 39 University Housing Resident Assistants who were hired the term following the concurrent validity study.

Procedure

A critical incident job analysis had been previously conducted of the University Housing Resident Assistant position (Aamodt, Keller, Crawford, and Kimbrough, 1981). The results of this job analysis indicated that there were 13 important dimensions of Resident Assistant performance. Tests were selected that proportedly measured each of the dimensions found in the job analysis and were then administered to the incumbent Resident Assistants. The test results, in turn, were correlated with performance ratings provided by the respective Head Residents for each of the dimensions. All ratings were made on 9-point graphic rating scales. The tests that correlated highest with each of the dimension ratings were entered into a stepwise regression equation using the Statistical Analysis System (SAS) in order to predict an overall performance rating of each Resident Assistant that was obtained by having the Head Residents make overall ratings on a 9-point graphic rating scale.

The results indicated that the responsibility and compliance scales of the Trait Evaluation Index were the only variables that reached a significant ($p < .05$) level of predictability. The two scales were then weighted in each of the four following ways:

Rank Order Weights—All Resident Assistants in the concurrent validity study were asked to rank order the importance of the dimensions that were found in the job analysis. Mean ranks were then computed for each dimension.

Unit Weights—The standard scores of the two predictors were each given weights of unity.

Critical Incident Weights—The job analysis mentioned previously had been conducted by having incumbents write critical incidents of Resident Assistant performance. Thirteen categories emerged from this procedure and the number of incidents that had been sorted into each of the categories was used as a measure of the importance of that dimension (Glickman and Vallance, 1958). Thus, in the present

TABLE 1
Weights Derived Within Each Weighting Method

	Performance Dimension	
Weighting Method	Responsibility	Compliance
Rank Order	4.63	5.09
Critical Incident	35.00	18.00
Unit Weights	1.00	1.00
Regression	.38	.62

study, the two predictors (responsibility and compliance) were weighted by the number of critical incidents that were sorted into each of the respective categories.

Regression Weights—The weights that were derived from the stepwise regression procedure mentioned earlier were utilized for this method of weighting.

The two predictors were then combined using each of the four weighting methods to form a composite score. Table 1 shows the various weights that were assigned to the two predictors.

Table 2 shows the correlations among the four methods of weighting for both the concurrent and cross-validation samples. As indicated in Table 2, all methods are highly correlated with eight of the twelve coefficients reaching levels of .95 or greater. All values are significant at the .01 level. These results tend to support and extend the findings of Fralicx and Raju (1982). It is significant to note that this study differed from the Fralicx and Raju (1982) study in many ways (predictors instead of criteria, different weighting methods, fewer dimensions, cross-validation, and a different sample) yet the results yielded the same basic conclusion.

Table 3 shows the validity coefficients for both the concurrent and cross-validation groups. Severe validity shrinkage is apparant

TABLE 2
Intercorrelations Among the Four Weighted Composites

	Weighting Method			
Sample/Weighting Method	Rank	CIT	Unit	Regression
Concurrent Sample				
Rank Order	1.00	.93	.99	.98
Critical Incidents		1.00	.95	.84
Unit Weights			1.00	.97
Regression				1.00
Cross-validation Sample				
Rank Order	1.00	.94	.99	.98
Critical Incidents		1.00	.96	.86
Unit Weights			1.00	.99
Regression				1.00

TABLE 3
Validity Coefficients Resulting from the Weighting Methods

Weighting Method	Validation Type Concurrent	Cross
Critical Incident	.38	.23
Rank Order	.46	.19
Unit Weights	.45	.20
Regression Weights	.48	.16

regardless of the method used to weight the data. However, tests for significant differences between any pair of validity coefficients indicates that none of the weighting methods provided anymore protection against validity shrinkage than did another.

It appears that when the results of this study are combined with those of Fralicx and Raju (1982), the method used to weight multiple predictors or criteria in the forming of a composite score is not as important as was once believed. This conclusion is consistent with the thoughts of Wainer (1976) who theorized that equal weights are as accurate and perhaps more robust than regression weights due to the effects of outliers and the overfitting of the model to the original sample that occur with regression.

In light of the research by Schmidt (1971), it was somewhat surprising that in the current study, no significant differences occured among the methods in terms of resistance to validity shrinkage. However, even though the differences were not significant, more shrinkage did occur when regression weights were used than when any other method was used. Thus, regression weights may optimize the validity coefficients in a concurrent design (even though they were not significantly better than the other methods in this concurrent validity study), but are no more effective down the line than are the other methods. More research is necessary to further clarify the importance of different methods of weighting predictors.

REFERENCES

Aamodt, M. G., Keller, R. J., Crawford, K. J., and Kimbrough, W. W. (1981). A critical incident job analysis of the University Housing Resident Assistant position. *Psychological Reports*, 49, 983–986.

Einhorn, H. J. and Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 13, 171–192.

Fralicx, R. D. and Raju, N. S. (1982). A comparison of five methods for combining multiple criteria into a single composite. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 42, 823–827.

Glickman, A. S. and Vallance, T. R. (1958). Curriculum assessment with critical incidents. *Journal of Applied Psychology*, 42, 329–335.

Lawshe, C. H. (1969). Statistical theory and practice in applied psychology. *Personnel Psychology*, 22, 117–124.

McCormick, E. J. and Ilgen, D. R. (1980). *Industrial Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Nelson, A. R. (1966). *Examiner's Manual: Trait Evaluation Index*. New Rochelle, NY: Martin M. Bruce Publishers.

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 31, 699–714.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83, 213–217.