Florida Journal of Educational Research Fall 1986, Vol. 28, No. 1, Pp. 5-23

Assessment of School Merit with Multiple Regression: Methods and Critique

Richard L. Tate
Florida State University

ABSTRACT. Regression-based adjustment of student outcomes for the assessment of the merit of schools is considered. First, the basics of causal modeling and multiple regression are briefly reviewed. Then, two common regression-based adjustment procedures are described, pointing out that the validity of the final assessments depends on (a) the degree to which the assumed adjustment model accurately reflects the actual causal processes of schooling in the district and (b) the validity and reliability of the measurement of all necessary variables. In the final section, it is argued that assessment of school merit should not be based solely on regression-based adjustment of student outcomes because state-of-the-art knowledge of causal processes and measurement skills are not adequate to ensure reasonable accuracy. A simple example is used to illustrate the possible severity of bias introduced by a single "specification error" in the model. It is suggested, however, that regression-based procedures might play other, less central, roles in the assessment of school merit.

Serious complications associated with the direct assessment of school merit based on identification of

---

This article is based on a report supported by a grant from the Florida Department of Education. The views expressed herein do not necessarily reflect those of this agency.

meritorious practices and policies often lead to proposals for an indirect approach focusing on student outcomes which are presumed to be influenced, in part, by school merit. In principle, such a procedure would be based on any valued outcomes, including student attitudes, motivation, citizenship, and study habits; but general academic achievement has been the most common outcome proposed and will be the one assumed for illustration.

It is, of course, not legitimate to compare schools directly on, say, student results from a standardized achievement test, even if the test happens to accurately reflect instructional objectives across all schools in a district. Schools differ on many other factors besides merit which are also determinants of student achievement; some of the factors most commonly identified are student ability and motivation, family background variables, and school resources and facilities which are not under direct control of school personnel. Thus, for an indirect assessment of school merit, it is necessary to "adjust" observed achievement differences among schools for any differences on other important unrelated determinants of achievement. Multiple regression is the analysis procedure most commonly proposed and used for this purpose.

In this article, some basic concepts of causal modeling and multiple regression will be reviewed first. Then, two common ways of using multiple regression for the indirect assessment of merit will be described, emphasizing the importance of the assumed causal model in developing the analysis model and identifying the conditions which must be satisfied for valid assessment. A critique of these two approaches is offered in the last section of the chapter.

Causal Modeling and Multiple Regression

Structural models. Actual computations in the indirect assessment of school merit often involve only the technique of multiple regression, which will be briefly described. Any specific regression equation, however, implies a theory of the causal processes in schooling which may or may not be reasonable.

Therefore, the critical question of the validity of the resulting assessment must be addressed with causal modeling.

Begin by assuming, for illustration, a model of reality in which the average student achievement at the end of the current year (denoted CA_j for the jth school in the district) is directly determined by only three variables merit of the school during the current year (MER_j), average achievement of the same students at the end of the previous year (PA_j) and average student motivation during the current year (MOT_j). This assumed model implies that all other determinants of current achievement are indirect causes, i.e., they influence current achievement only through their effects on the three direct effects identified above. (The causal model to be used here is almost certainly a drastic oversimplification of reality but will suffice for illustrating important points.) This causal model, called the structural model, is represented by the "path diagram" in panel A of Figure 1. The straight arrows represent direct causal effects, each arrow or path being labeled with a structural coefficient indicating the direction and magnitude of the effect. Note that the variables in the diagram have been numbered (the missing number 3 will be introduced later); the subscripts of the structural coefficients are based on these numbers, e.g. the coefficient P_{52} is the direct effect of previous achievement (variable 2) on current achievement (variable 5). The curved arrows represent unanalyzed correlations among the MER, PA and MOT variables.

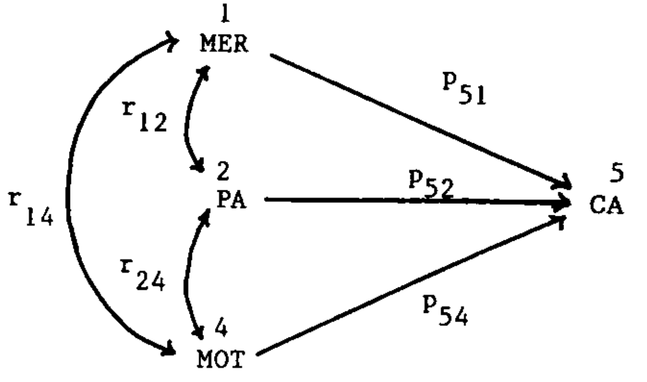
The structural model is represented in equation form by:

$$CA_j = P_{51}MER_j + P_{52}PA_j + P_{54}MOT_j.$$

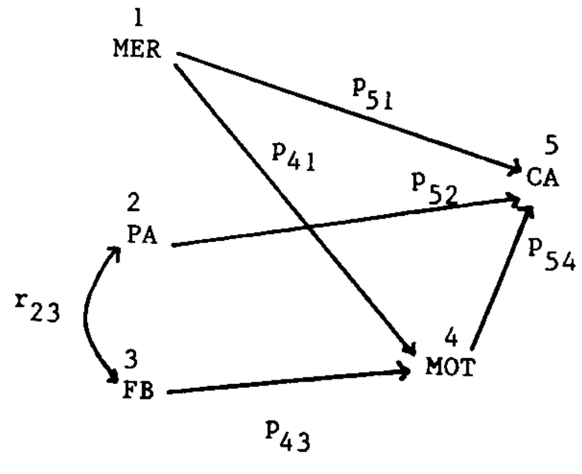
All variables are assumed to be in standardized form. In the multiple regression literature, CA is usually called the dependent variable and MER, PA, and MOT are the independent variables, while in the causal modeling literature, CA is the endogenous variable, or the variable explained by the model; the others are exogenous variables, or variables not explained by the model. The latter terminology is more appropriate

Figure 1

Assumed Structural Models for Illustration



(A) Assumed model for direct effects (MER = merit, PA = previous achievement, MOT = motivation, and CA = current achievement)



(B) Complete model: Direct and indirect effects (adding FB = family background) for the purposes here.

The structural coefficients in the model are assumed here to be $P_{51} = 0.292$, $P_{52} = 0.5$, $P_{54} = 0.5$. Thus, if a given school is one standard deviation above the mean on merit (MER = 1), a half standard deviation below the mean on previous achievement (PA = -0.5), and one and a half standard deviations above the mean on current motivation (MOT = 1.5), the current achievement will be, in assumed reality,

$$CA = (.292)(1) + (.5)(-0.5) + (.5)(1.5) = .79,$$

or 0.79 standard deviations above the mean for all schools on CA. The individual structural coefficients also, as indicated above, give the direction and magnitude of each direct effect; e.g., the coefficient $p_{51} = 0.292$ indicates that an increase of one standard deviation on MER, holding constant PA and MOT, would cause a 0.292 standard deviation increase in CA.

Multiple regression. The structural coefficients can be estimated with multiple regression, a technique which permits the prediction of a single endogenous variable with multiple exogenous variables. Using the standardized variables already introduced, the prediction equation for current achievement for the jth school would be stated

$$\hat{CA}_j = P_{51}MER_j + P_{52}PA_j + P_{54}MOT_j.$$

(Temporarily assume that there is a direct measure of merit, MER, to illustrate some basic concepts, before the situation in which there is no such direct measure is addressed.) The "hat" over CA indicates the predicted value and the coefficients are usually called standardized partial regression coefficients. In any population of schools, there will be a prediction error, E_j, for each school defined as the difference between the observed and predicted values of the endogenous variable, i.e., $E_j = CA_j - \hat{CA}_j$. Multiple regression computes the regression coefficients by finding those values called "least squares" estimates which minimize the sum of the squared residuals for all observations in the sample or population. This is equivalent to finding the coefficients which maximize the bivariate correlation between CA and $\hat{CA}$. The overall strength of the relationship between CA and the multiple independent variables is indicated by the coefficient of determination, R^2, defined as the proportion of CA variability (expressed in terms of sums of squares) which is explained by the model. The R^2 index is also the square of the bivariate correlation between CA and $\hat{CA}$. The standard deviation of the residuals, called the standard error of estimate (SE), reflects the unexplained variability.

It should be noted that multiple regression is most typically applied to data from a sample which has been randomly selected from a population of interest. For the problem of interest here, however, data from the entire population can usually be assumed, that is, data would be collected from all schools in the district of interest. Therefore, it is not necessary to consider here procedures of statistical inference (hypothesis testing and interval estimation) associated with multiple regression. Instead, multiple regression is used only to compute population parameters.

Multiple regression can be used for an unbiased computation of the structural coefficients (i.e., the true causal effects), given the validity of the following assumptions:

A. The analysis model is correctly specified. This is simply the assumption that the analysis model correctly reflects the actual causal processes. There are three important elements of this assumption

1. Variables having direct effects on the endogen--ous variable which are correlated with other direct effects can not be omitted. (Any direct effect which is uncorrelated with all other direct effects may, however, be omitted from the equation without biasing results.)

2. There is no reciprocal causation (i.e., causal feedback) in which the endogenous variable in turn influences one or more of it's direct determinants. (A model with no reciprocal causation is called a "recursive" model. The coefficients in "nonrecursive" models with causal feedback cannot be computed directly with multiple regression, but other computational procedures are available.)

3. The correct functional form of the model has been specified. For example, if in reality there is an interaction between two of the direct effects, the analysis model must also contain this interaction.

B. All variables which have direct effects on the endogenous variable are measured exactly, i.e., there is no measurement error. (No bias is introduced by measurement error for the endogenous variable.)

If all of these assumptions are valid, multiple regression provides unbased estimates of the structural coefficients. To illustrate this assume that the model in panel A of Figure 1 accurately reflects the true causal processes for all schools in the district. Assume further that the analyst decides to use the same variables in the analysis model. If perfectly reliable measures on all four variables are collected for all schools, it can be shown that the bivariate correlations among the variables must be those shown in Table 1. When these correlations are used as input for an analysis model reflecting the true structural model, the resulting regression coefficients are identical with the corresponding structural coefficients. (The associated R^2 is equal to 1 since complete determination and perfectly reliable measurement has been assumed for this illustration.)

If any of the assumptions are violated, then some degree of bias is present in the computed structural coefficients. That is, given population data, the computed multiple regression coefficients are not equal to the true structural coefficients. Of course, violation of assumptions and associated bias is, in practice, a question of degree. Slight violations of certain assumptions may result in bias which is of no practical importance.

Although empirical data can be very useful in considering the validity of some of the above assumptions (e.g., correct functional form and reliable measurement), there is unfortunately no empirical test of the assumption that all appropriate variables have been included in the model. The assumption of correct specification must rest on the soundness of the reasoning which resulted in the model specification.

<u>Indirect and total effects</u>. If primary interest is

Table 1

Assumed Coefficients for Models in Figure 1

Assumed Structural Coefficients	Implied Correlations	MER	PA	FB	MOT	CA
$r_{23} = .6$	MER	1.0				
$p_{41} = .6$	PA	0	1.0			
$p_{43} = .8$	FB	0	.6	1.0		
$p_{51} = .292$	MOT	.6	.48	.8	1.0	
$p_{52} = .5$						
$p_{54} = .5$	CA	.592	.74	.7	.915	1.0

in the causal effect of school merit on current achievement, the assumed structural model in panel A of Figure 1 tells only part of the story. Thus far, A only direct causal effects have been considered, and it is very likely that merit also has indirect effects on current achievement. For example, given the assumed direct effects in the current model, it is reasonable to suppose that merit is also one of the direct determinants of current motivation. This assumed causal link, represented by the arrow from MER to MOT in panel B of Figure 1, is set equal to 0.6 for this example. There are now two causal paths from MER to CA, the direct effect with strength 0.292 and the indirect effect represented by the path from MER to MOT to CA. The strength of this indirect effect is equal to the product of the coefficients for the two associated paths (i.e., $P_{41}P_{54} = (.6)(.5) = .3$). The total causal effect of MER on CA, then, is the sum of the direct and indirect effects or $0.292 + 0.3 = 0.592$.

Since current motivation is now an endogenous variable, i.e., a variable being explained by the model, all variables directly affecting it must also be included for correct specification. Assume that a "family background" variable (labeled FB) is the only other direct determinant of MOT with an effect strength of 0.8. The resulting complete causal model is shown in Panel B of Figure 1.

One last important comment about the final model is necessary. Note that there is a curved arrow between PA and FB, representing the unanalyzed correlation between these two variables. The absence of curved arrows between MER and PA and between MER and FB signify zero correlations between school merit and the other two exogenous variables of previous achievement and family background. This is consistent with the typical definition of merit as something like "productive effort of the school, given student and family characteristics." In other words, merit is often viewed not as the absolute quality of instruction, but rather as that component of quality and effort which is independent of factors not under the school's control.

The complete process of model specification, estimation of the coefficients, and description of direct, indirect, and total effects is called path analysis or causal modeling. When there is no reciprocal causation, multiple regression can be used to compute the model coefficients by regressing each endogenous variable on only the direct effects of that variable. For the example, MOT would be regressed on MER and FB, producing the P_{41} and P_{43} coefficients, and CA would be regressed on MER, PA, and MOT (as illustrated earlier) for determination of the P_{51}, P_{52} and P_{54} coefficients. The same assumptions identified earlier must now be valid for the equation for each endogenous variable to produce unbiased computations. For example, if the family background (FB) variable is not included, the computation of the other direct effect (P_{41}) of MOT is biased, producing a biased indirect effect of MER on MOT.

When only the total effect of MER on CA is of interest and it is not important to know what portion is direct or indirect, an alternative regression analysis can be used. An equation regressing current achievement on the exogenous variables of the full model is called the "reduced model" and produces the total effects of each of the exogenous variables on CA. That is, in the regression equation,

$$\hat{CA} = P_{51}{*}MER + P_{52}{*}PA + P_{53}{*}FB,$$

the coefficient $P_{51}{*}$ represents the direct plus indirect effect of MER on CA ($P_{51}{*} = P_{51} + P_{41}P_{54}$). Input of the appropriate correlations from Table 1 into a regression program for this model results in $P_{51}{*} = 0.592$ which is identical to the total effect found earlier by computing the direct and indirect effects separately.

Finally, it was mentioned earlier that omission of a direct effect which is uncorrelated with all of the other direct effects does not bias the computation of the remaining coefficients. The same is true for the reduced model. Merit is often defined to be uncorrelated with PA and FB; given such a definition, if MER were omitted from the reduced analysis model, the analysis would still result in the correct total effects of PA and FB on CA. This feature is central to the approach discussed next.

Merit as Residuals

For purpose of exposition, discussion to this point has assumed that a valid measure of school merit is available. Of course, the primary concern here is the indirect assessment of merit when such a direct measure is not available. In the "merit as residual" approach, the analyst wants to specify an analysis model which exactly reflects reality except for the unmeasured merit variable. Assume for illustration that the true causal processes are accurately reflected by the reduced structural model presented previously, i.e.,

$$CA_j = P_{51}*MER_j + P_{52}*PA_j + P_{53}*FB_j,$$

in which CA is completely explained and the coefficients are the total effects. For the "merit as residual" approach, the appropriate prediction equation would therefore be

$$\hat{CA}_j = P_{52}*PA_j + P_{53}*FB_j.$$

Since the only variable which has been left out is, by definition, uncorrelated with the other variables in the equation, the computation of the remaining coefficients is unbiased.

With perfectly reliable measurement, the residual in the prediction equation exactly reflects the effect of merit, i.e.,

$$\begin{aligned} E_j &= CA_j - \hat{CA}_j \\ &= P_{51}*MER_j. \end{aligned}$$

Since the residual is proportional to merit, the residuals for all schools could, in principle, be computed and compared to indicate relative merit. The standard error of estimate, SE, can be used to assist in describing the relative merit of any single school. To illustrate, the regression of CA on PA and FB results in the prediction equation CA = (.5)PA + (.4)FB, with an associated SE of 0.592. Suppose one school has CA = 0.3, PA = 1.0 and FB = 0.5; the predicted achievement would be $\hat{CA}$ = (.5)(1.0) + (.4)(.5) = 0.7, and the residual $\hat{E}$ = 0.3 - 0.7 = -0.4. Thus, the school is 0.4/.592 = 0.69 standard deviations below the mean of the merit distribution. If the distribution is approximately normal, reference to a normal curve table indicates the school is at about the 25th percentile in the merit distribution. Application of the same process to all other schools in the district results in a ranking of the schools.

The conditions which must be satisfied in order to obtain valid assessments of school merit with the "merit as residual" approach follow the causal modeling assumptions identified in the previous section. That is, the analysis model must be correctly specified (i.e., accurately reflect the true casual processes), and all variables in this model must be measured validly and reliably. These assumptions are important in practice, as will be illustrated in the final section.

Merit as Adjusted School Effects

A second common approach, also based on multiple regression, shifts the focus from the school level to the student level. The concern now is with the causal determinants of the current achievement for the ith student in the jth school (labeled CA_{ij}). Suppose the true reduced structural model is again based on the full model shown in panel B of Figure 1, except the determinants PA, FB, and MOT are now measured at the student level. The reduced structural equation is

$$CA_{ij} = b_{51}MER_j + b_{52}PA_{ij} + b_{53}FB_{ij}.$$

It is now assumed that each student-level variable is standardized with respect to the student population in the entire district; different coefficient symbols have been used because the effect of, say, PA_j on CA_j (school level) is conceptually different from the effect of PA_{ij} on CA_{ij} (student level). Note that the value of MER is constant for all students in the same school, j, so the model can also be expressed as

$$CA_{ij} = b_{oj} + b_{52}PA_{ij} + b_{53}FB_{ij}.$$

This model now has a constant or intercept term, b_{oj}, which varies only from school to school and reflects the effect of school merit on CA_{ij}; i.e.,

$$b_{oj} = b_{51}MER_j.$$

Thus, the difference in merit between any two schools is proportional to the difference of the constants in the associated equations, i.e., the difference in merit between two schools, j and k, is proportional to the difference $b_{oj} - b_{ok}$. Another way to think of the same difference defines the "adjusted mean achievement" of a school as the predicted CA_{ij} for PA_{ij} and FB_{ij} equal to zero (i.e., for values of previous achievement and family background equal to the respective grand means for the district). The difference in adjusted means for schools j and k (the difference in achievement controlling for other school differences) is also equal to the intercept difference and proportional to the merit difference of interest.

Given student data for all schools in the district, an analysis of covariance (ANCOVA), a special case of multiple regression, can be used to compute the adjusted mean difference for any two schools. For the assumed reality, the analysis model should include school as a categorical variable in addition to the PA and FB variables. A categorical variable with k levels can be represented with a set of k-1 coded "dummy" variables. If, for example, there are 30 arbitrarily numbered schools in the district, then the first variable, say X_1, in the required set of 29 variables would be coded "1" for all students in the first school and "0" for all other students; the second variable in the set, X_2, would be coded "1" for all students in the second school and "0" for all other students, etc. The regression coefficients b_1, b_2, etc. for the variables in the set then provide the desired adjusted mean differences which, if the analysis model is correct, are proportional to the merit differences. Specifically, the coefficient for the jth variable in the set of dummy variables is the difference between the adjusted mean for the jth school and that for the last or "control" school. Consideration of all of the 29 coefficients will provide a ranking of all schools relative to the control school. If, for example, the coefficients of say, schools number 8, 13, 24 and 27 are 0.1, -0.3, 0.2, and -0.1, respectively, the relative merit ranking of these schools plus the control school is, from highest to lowest, school 24, school 8, control school, school 27 and school 13.

The same conditions of correct specification and valid and reliable measurement discussed previously must also be satisfied here for valid assessment. It should be noted specifically that the model must be correctly specified at both levels, student and school.

Critique/Recommendation

To this point, two common indirect approaches to the assessment of school merit and the associated conditions which must be satisfied to produce valid assessments have been described. Now consider the feasibility of sufficiently meeting in practice the conditions required for reasonably valid merit assessments.

This writer believes that the current research literature in a number of different areas suggests clearly that any assessment of school merit based solely on regression-based adjustments of student outcomes should be viewed with extreme skepticism. Consider first the assumption that valid assessment requires an analysis model which accurately reflects the true causal processes. Research literature in educational psychology, instructional development, classroom processes, and school effectiveness all indicate that, despite some progress over the years, our current theories and models of the various aspects of schooling are still relatively primitive and incomplete compared to the recognized complexity of the phenomenon. Models in the same area often differ widely and a listing of all of the variables suggested in the various admittedly incomplete models is strikingly long. Moreover, researchers are just beginning to identify some of the complex interactions which exist among the determinants of student achievement. In other words, there is no sense that there is or will be in the foreseeable future a confident consensus on the true causal processes of schooling in different settings and circumstances. This is, of course, not to say that an individual knowledgeable in the appropriate substantive areas, the schools in the district of interest, and causal modeling could not develop a reasonable and sophisticated model. The point is simply that someone else just as knowledgeable may develop a very different model, with different implications for the assessment of merit, which seems just as reasonable. And, as mentioned earlier, current research methodology offers no way to prove that one of the models better reflects reality.

The importance of the correct specification assumption can be illustrated by showing how a simple specification error can produce unacceptable bias in the assessment of merit. Assume that the "merit as residual" approach is being used and that the model in panel B of Figure 1 accurately reflects the true causal processes in a school district. As discussed in a previous section, use of the correctly specified model (omitting only the unknown merit variable) would result in computed residuals which exactly reflect the actual merit of each school. Suppose, however, the individual responsible for developing the assessment procedure believes that the only important determinants of current achievement (CA) are merit and past achievement (PA), and regresses CA on PA only, leaving out the family background variable (FB). The resulting bias is described in Table 2. Each column heading indicates an actual merit rating for a school (expressed as a percentile rank), while each row is for a different combination of previous achievement and family background (PA/FB). (Only PA/FB combinations which are possible given a moderately strong correlation between PA and FB are considered.) Each table entry gives the computed merit ranking using the incorrectly specified model for a specific PA/FB and actual ranking combination. Bias is reflected by differences between actual and computed merit rankings within each column.

Table 2

Computed Versus Actual Merit Rankings[a]

Previous Achievement/ Family Background	Actual Merit Ranking 02	16	50	84	98
0/0	04	19	50	81	96
2/1	03	16	45	78	95
1/2	18	48	80	96	100
-1/1	21	53	83	97	100
-2/-1	05	22	55	84	97
-1/-2	00	04	20	52	82
1/-1	00	03	17	47	79

[a]Column headings and table entries are merit percentile rankings. Previous achievement and family background are expressed in terms of z scores.

There is some degree of bias for most of the conditions represented in Table 2. The bias is most severe for the PA/FB combinations of 1/2, -1/1, -1/-2, and 1/-1, with the difference between actual and computed merit reaching 30 percentile points and more. For example, when a school has an average previous achievement which is one standard deviation above the district mean, an average family background one standard deviation below the mean (i.e., PA/FB = 1/-1), and the school's actual merit ranking is 84, use of the incorrectly specified model would result in a computed ranking of 47. That is, the school would be identified as being in only the second quartile, when in reality it is in the fourth or top quartile. Thus, the omission of only one important variable has resulted in an unacceptable bias for certain types of schools. The direction of the bias for the conditions with the worst bias is determined, for this example, by the sign of the family background (FB) variable. If FB is positive, computed merit is larger than the actual (positive bias), while the reverse is true for negative FB values.

A second important condition for the valid assessment of merit with regression-based approaches is the valid and reliable measurement of the variables in the model. Again, even though advances in measurement theory and practice have led to confidence in the assessment of certain types of variables (e.g., student achievement), the literature indicates serious unresolved difficulties and lack of consensus in measuring more abstract constructs like motivation and critical aspects of family background which may be in any correctly specified model. For example, there is no reason to believe that the easily measured "percent of students in a school receiving free lunch" would be a valid measure of the aspect of family background which is the actual causal determinant of student achievement. Thus, even if, by insight or luck, the analysis model is correctly specified, the degree of invalidity and unreliability associated with current state - of - the - art measurement may still produce acceptable bias.

Given these fundamental concerns, it is clear that the assessment of school merit should not be based solely on regression-based adjustment of student outcomes. There are, however, several ways in which these procedures may still be of value. First, any approach to the assessment of school merit must be based on a theory of the schooling process in the district or state. This theory is explicit in the indirect approach discussed here and implicit in any direct assessment based on identification of meritorious practices and procedures. The causal modeling procedures briefly described offer a way to empirically test different developing theories in order to refine the theoretical basis of any assessment. A second, perhaps somewhat remote, possibility is that a district may develop an assessment procedure, based perhaps on extensive observation in all schools of the district, in which there is full confidence on the part of all concerned parties. This accepted valid measure could then be used as a criterion to validate a regression-based assessment. If the results of the two approaches are consistent for all types of schools over several years, use of the regression approach rather than the (probably) more expensive and inconvenient observation procedure would be justified.

Finally, perhaps the most likely possibility is that there are currently serious flaws associated with any single proposed approach to merit assessment. In this case, it seems likely that the greatest confidence in the fairness of any merit assessment would result when the assessment system is based on several different approaches, including adjusted student outcome comparisons, and is tempered by informed judgment. For example, if similar rankings for a given school emerge from both adjusted student outcomes and the direct observation of school process, the two results could be averaged and used without modification. On the other hand, if there is a serious inconsistency between these two different types of results for a given school, a group of individuals representing all concerned parties could consider all evidence, making whatever adjustments are appropriate given knowledge of the school and district, and make a final assessment.

References

Asher, H. B. (1983). Causal modeling (second edition). Beverly Hills, CA: Sage Publications.
Duncan, O. D. (1975). Introduction to structural equation models. New York: Academic Press.
Heise, D. R. (1975). Causal analysis. New York: Wiley.
Kenny, D. A. (1979). Correlation and causality. New York: Wiley.

AUTHOR
RICHARD L. TATE, Associate Professor, Department of Educational Research, Development, and Foundations, Florida State University, Tallahassee, Florida 32306