Educational Considerations

Volume 39
Number 1 *Class Size and Student Achievement*

Article 5

9-1-2011

A Practical Method of Policy Analysis by Estimating Effect Size

James L. Phelps

A Practical Method of Policy Analysis by Estimating Effect Size

James L. Phelps

The previous articles on class size and other productivity research paint a complex and confusing picture of the relationship between policy variables and student achievement. Missing is a conceptual scheme capable of combining the seemingly unrelated research and dissimilar estimates of effect size into a unified structure for policy analysis and decision making. This article builds a rationale for a unifying structure and consistent method of estimating effect size.

Forrester (1980), in his work on system dynamics, offers pertinent ideas. He stressed the importance of constructing a comprehensive operating structure to better understand an organization's complexity and its behavior in response to policies. By structure, he meant all the diverse elements of the organization, including their specific responsibilities and, most importantly how the elements related to one another in some quantifiable manner. Within the identified operating structure, policy decisions were made to directly influence changes in behavior in specific elements of the organization. Those same policies also indirectly influenced other elements of the organization because the elements were interrelated. Quantifying these elements and their interrelationships within a unified scheme is essential to the workings of system dynamics. This model relies on a set of parameters to simulate organizational behavior in response to various policy options. The purpose of the model is to predict how policy changes will influence organizational behavior which, in turn, will achieve the desired outcomes.

Another representation of the organization is what economists call a production function. The outcomes (outputs) of the organization are the byproducts of the resources (inputs) and the processes used to convert the resources into outcomes. Using this framework, the educational outcomes are achievement measures; the resources are services and materials purchased, e.g., staffing; and the processes include the curriculum, instructional program, and home activities, for example. In most production function studies, however, little attention is paid to the process variables largely because of the lack of data and a meaningful method of assimilation. When interpreting the results, primary attention is directed to the linear weights, or regression coefficients. Less attention is paid to the statistics describing the explained variance (R^2) and the residual. These statistics provide a different approach to a unified structure and method of estimating effect size. The main purpose of the production function is to estimate the parameters of a small set of relationships and make probability inferences. Most econometric studies focus on class size or some other narrow aspect of education rather than the entirety of school activities. As a result, econometrics has substantial limitations in simulating organizational behavior for multiple goals and policy options.

A desirable paradigm would combine features from both system dynamics and econometric modeling. A semantic clarification is in order. Here, I am referring to a paradigm as a model, and a model as a hypothetical formulation used in analyzing or explaining something. In the context of this article, the paradigm is the formulation of a unified school structure including what Kuhn (1970) labeled theory, laws, application, and instrumentation. The model is the mathematical representation of the paradigm, or the laws, application, and instrumentation components of the paradigm. Based on these concepts, the immediate task is to identify the resource and process elements of the educational organization and quantify their relationships with the outcomes, all under some unifying scheme or structure—in other words a paradigm.

This article develops a policy analysis paradigm by combining the various estimates of effect sizes into a coherent structure with a consistent method of measurement; and building a rational and analytical method to accommodate the effect ceiling and effectiveness components. The final product is a suggested analytic structure, a list of characteristics associated with the method of measuring effect size, and a list of assumptions underlying the policy analysis paradigm. Finally, there is a compilation of estimated effect sizes.

What makes this paradigm "sufficiently unprecedented," to use Kuhn's phrase, is the method of estimating effect size permitting the principles of system dynamics to be incorporated into a method of policy analysis. The effect sizes, when coupled with the incremental cost of the policy options, provide policymakers with a model to evaluate the potential achievement gains based on various combinations of alternatives (Kuhn's application and instrumentation). This final stage of the paradigm addresses three overarching questions:

- Under what circumstances might lowering class size be effective?
- What are the competing resource and process policies for improving achievement?
- How do policymakers decide what is the most effective and efficient course to follow?

The first section in this article reviews the conceptual issues related to the relationship between class size and achievement, as follows: Measurement of the concentration of teachers and students; collinearity among the data variables; influence of socioeconomic status (SES) as an intervening variable; and modeling the relationship between achievement and policy options. Section two provides estimates of effect size from a Minnesota data set, utilizing different statistical methods to illustrate the various methods available to measure the magnitude of effect size. It highlights the difficulties in measuring effect size and demonstrates a method to

---

James L. Phelps holds a Ph.D. from the University of Michigan in Educational Administration. He served as Special Assistant to Governor William Milliken of Michigan and Deputy Superintendent in the Michigan Department of Education. Active in the American Education Finance Association, he served on the Board of Directors and as President. Since retirement, he spends a great deal of time devoted to music, composing and arranging, playing string bass in orchestras and chamber groups, as well as singing in two choirs. He resides with his wife, Julie, in East Lansing, Michigan.

place the various estimates into a unified structure. These estimates are compared with those from the studies reviewed in the previous article. Section three summarizes the material presented and states the assumptions guiding a policy analysis model.

Conceptual Issues

Measurement of the Concentration of Teachers and Students

The method of measuring the concentration of teachers and students has cost implications as demonstrated by this example: The additional cost of reducing the class size from 20 to 19. This raises a concept from physics known as the quantum jump, or the energy required for an electron to jump from one energy state to another. (The energy comes only in well-defined packets. Such is the case with class size.) If there are 60 students in a particular grade, then class size is determined by the number of teachers assigned to that grade. The number of teachers is the quantum number, not the number of students.[1] With 1 teacher, the class size is 60; with 2, the class size is 30; with 3, it is 20; and, with 4, it is 15. In other words, there is no possible way of reducing class size from 20 to 19. In order to lower the class size below 20, the only policy alternative is to add one additional teacher and pay the costs to reduce the class size from 20 to 15. Therefore, the appropriate policy-oriented class size measure is the teacher/pupil ratio.

Collinearity among Explanatory Variables

There is no perfect way to measure effect size. First, there is always a degree of measurement error. Second, in most cases, explanatory variables are intercorrelated. For example, in the case of two explanatory variables, the influence (proportion of variance explained, or R^2) is divided into segments: The unique influence of each variable and the common influence among the variables. There is no unequivocal way to partition the common influence into the unique influence of both variables. The regression process attributes the common influence to the variable with the highest correlation with the achievement variable, most likely SES. Therefore, the variable of policy interest, the teacher/pupil ratio, is allocated the remaining portion of the explained variance and, as a result, a lower weighting. When there are two variables, the compromise is to estimate the maximum effect size (with the common variance) and minimum effect size (without the common variance) for the policy variable and select the appropriate value on other grounds. This same principle applies to the many instructional variables identified by Walberg (1984)[2] and explains why his estimated effect sizes could not be added—they were correlated! When there are more than two variables, it is desirable to combine the effect sizes into a cluster, or factor, containing all the unique and common variance (Phelps, 2009).

Influence of Socioeconomic Status (SES) as an Intervening Variable

Over the years, federal and state governments have provided additional funds to low performing schools. These are determined in a number of ways, usually by achievement scores or SES. Schools receiving these funds often reduce their class size. As a result, it is likely that low-performing schools have lower class sizes. To adjust for this situation, a measure of SES in the analysis is critical. The inclusion of this intervening variable could materially change the magnitude of the relationship between achievement and the policy variable.[3]

Modeling the Relationship between Achievement and Policy Options

Regression is a statistical model to estimate the relationship between policy variables and achievement, but it has limitations pertaining to policy analysis. Because there can be but one regression equation, multiple achievement measures and variables with differing costs are not accommodated. There are other mathematical models addressing these shortcomings which are more helpful in evaluating policy alternatives. These models depend on simultaneous equations and nonlinear relationships between the outcome and the explanatory variables. There are substantial differences between nonlinear and linear models.

<u>Effect size for linear relationships: Constant slope</u>. Linear regression coefficients are the most frequent measure of effect size. The maximum effect size is estimated by regressing only the target variable with the achievement outcome either by the "b" weight or the standard regression coefficient expressed as Beta (β). The standard regression coefficient is more practical because it easily compares variables measured in differing metrics. SES could well be associated with class size, so it should be included as an intervening variable in the multiple regression equation to estimate the minimum.

<u>Effect size for nonlinear relationships: Changing slope</u>. It is highly unlikely that any policy variable will have a consistent, increasing or decreasing slope. Slight variations in the slope can be estimated by adding a squared term to the regression equation.[4] This does not provide either a theoretical or practical solution. There is, however, a theoretical sound and practical solution. This solution utilizes the amount of variance explained by the explanatory or policy variable in question, or the R^2.[5]

The R^2, when interpreted as the cumulative area under the normal curve, produces an S-shaped curve asymptotic at the top (maximum of 100th percentile) and bottom (minimum of zero percentile). If the R^2 is .5, then the S-shaped curve is reduced to the 75th percentile at the top and the 25th percentile at the bottom. As the R^2 approaches zero, the S-shaped curve approaches a line at the 50th percentile.

<u>Mathematical reason for the nonlinear relationship</u>. The difference between the linear and nonlinear interpretations can be demonstrated with a thought experiment using standard regression coefficients (β's). The regression equation states that the predicted outcome (measured in Z-scores) is equal to the sum of the β's times their respective Z-scores (and a percentile ranking can be calculated from any β and Z-score combination):

$$Y_{(z)} = \beta_1 Z_1 + \beta_2 Z_2 + \ldots \beta_n Z_n$$

The following calculations are for two hypothetical situations: (1) all Z-scores equal 1 (Z=1); and (2) all Z-scores equal 3 (Z=3). The variables are, SES, teacher/pupil ratio, instruction, and effectiveness. For each β*Z term in the equation, a percentile is calculated to measure the contribution to the overall change in performance. Assuming the starting point is the mean, the percentiles greater than .50 are calculated to determine the predicted gain. The percentile gains for the individual variables are then summed as indicated by the equation. (See Table 1.)

When each of the four variables is increased by 1-Z-score (from zero to 1), the increased percentile standing for all variables is .4236, or from .50 to .9236. When each variable is increased 3-Z-

Table 1
Calculation of Percentiles from Beta (β)

Variables	β*Z (Z=1)	Percentile	Percentile >.50	β*Z (Z=3)	Percentile	Percentile >.50
SES	0.8457	0.8011	0.3011	2.5371	0.9944	0.4944
Teacher/Pupil Ratio	0.0677	0.5270	0.0270	0.2031	0.5805	0.0805
Instruction	0.1200	0.5478	0.0478	0.3600	0.6406	0.1406
Effectiveness	0.1200	0.5478	0.0478	0.3600	0.6406	0.1406
Sum			0.4236			0.8560

Table 2
Calculation of Percentiles from R^2

Variables	R^2	$R^2/2$	Z-Score		
			- infinity	Z=0	+ infinity
SES	0.6827	0.3414	0.1587	0.5	0.8414
Teacher-Pupil Ratio	0.0280	0.0140	0.4860	0.5	0.5140
Instruction	0.0600	0.0300	0.4700	0.5	0.5300
Effectiveness	0.1400	0.0700	0.4300	0.5	0.5700
Subtotal	0.9107	0.4554	0.0447	0.5	0.9554
Error	0.0893	0.0447	0.4554	0.5	0.5447
Total	1.0000	0.5000	0.0000	0.5	1.0000

scores (from zero to three), the increased percentile standing is .8560. Because the starting point was the mean (.50), the increase brings the total to the impossible 1.356th percentile! Clearly, not all variables can be increased simultaneously. The β weights are partial regression coefficients and assume that all other variables stay fixed.

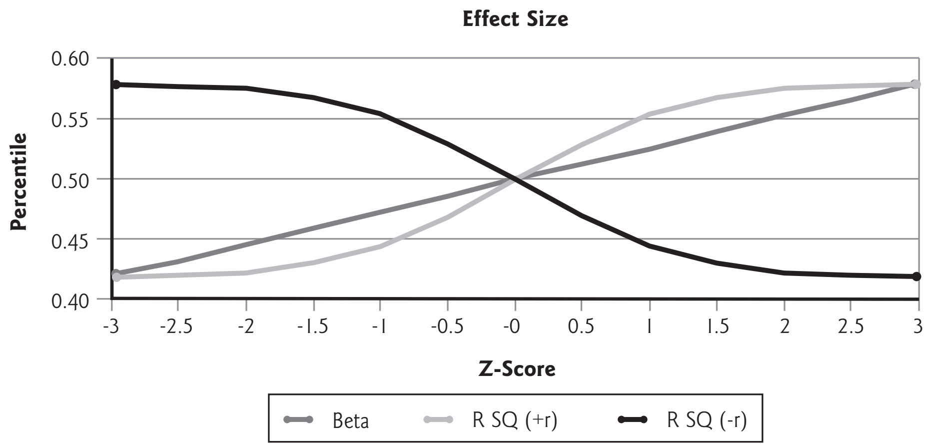
A second example uses the proportion of explained variance, or R^2, as the measure of effect size. To obtain the R^2, β is multiplied by the correlation coefficient: $R^2 = \beta_1 r_1$. The R^2 has four advantageous properties. First, the area under the normal curve is by definition equal to 1, so any point on the distribution can be defined as a percentile—the percent of observation below the point. Second, the highest point on the distribution is the 100th percentile and the lowest point is zero percentile. Third, the R^2 is the ratio between the outcome distribution and the explanatory distribution, so a percentile contribution to the outcome can be determined for any point on the explanatory distribution. Fourth, the mean (Z=0) on the explanatory variable will predict the mean of the outcome variable. Table 2 illustrates the percentile range (Z-score of +/- infinity) for each explanatory variable. One-half of the R^2 contribution is above the mean and one-half below. The R^2 values are listed with the minimum and maximum percentile levels. The contribution of the explanatory variables totals .4554 percentile points, ranging from .0447 to .9554.

Because the maximum R^2, including the error, for the variables is 1.00, no combination of variables, regardless of the Z-score can ever be higher than the 100th percentile or lower than zero percentile. In this case, there is no partial or fixed restriction as is the case with the regression β's. All variables are free to vary from the highest to the lowest Z-scores, accommodating the ceiling effect.

Figure 1 illustrates these different interpretations of effect size. The straight line represents the Beta coefficient between the extremes of Z-scores from zero to 3, but with all other variables fixed. The percentile ranking will continue to increase as the Z-score increases. The R^2 curve, the cumulative normal curve, is also between the extreme Z-scores, but with all other variables free to move. In contrast, the curve approaches a ceiling. The R^2 of any variable will have a negative sign if the regression coefficient is negative, as illustrated in Figure 1. The graph clearly depicts the difference between the unbounded character of the Beta coefficient and the ceiling character of the R^2.

Policy analysis differences between linear and nonlinear relationships. If a linear relationship is assumed with the β weight as the measure of effect size:

Figure 1
Representation of Beta (β) Weights and R^2 as Measures of Effect Size



Note: This graph is not to scale: Beta does not equal R^2.

- There can be only one best cost-effective policy, i.e., the variable with the largest standard regression coefficient (β) adjusted for cost.
- There is no reason to adopt anything but the most cost-effective policy option.
- The most cost-effective policy applies equally to all schools.
- There is never a point of diminishing returns.
- The linear relationships do not allow for an optimization process; i.e., finding the best combination of variables and costs to maximize the goals.
- Linear relationships are not an accurate representation of achievement production.

If a nonlinear relationship is assumed with R^2 as the measure of effect size and the residual as the measure of school effectiveness:[6]

- There is no one best cost-effective policy.
- The potential benefits will depend on the unique history of each school, i.e., their existing levels on all the policy variables, requiring unique policies for each school.
- When the benefit of a policy has reached a point of diminishing returns (high point on the S-shaped curve), a different policy with greater potential then becomes the preferred option.
- Nonlinear relationships are a more accurate representation of achievement production.

Recall the dilemma of Hedges, Laine, and Greenwald (1994)[7] as identified in the previous article; that is, spending money would improve achievement in every school even though no specific object for the funds was identified. Likewise, Glass and Smith (1978)[8] advocated lowering class size until there was one teacher for every pupil in order to achieve the maximum potential achievement. The list of instructional programs by Walberg also gave the same impression. In sum, if more funds, lower class size, and more instructional programs were provided, all schools would have unlimited success in raising achievement scores. No attention was paid to the ceiling imposed by achievement tests. No attention was paid to the uniqueness of every school setting. No attention was paid to the effective use of the resources or the quality of the instructional programs. Conclusions were based on the same mathematical model, the boundless regression line, which does not represent the realities of school operations.

If a different mathematical model is employed, one based on the statistical variance around the line, an entirely different notion emerges. Resources and instructional programs do make a difference, but the size of the difference is limited by the achievement test ceiling. The magnitude of these differences depends on the unique circumstances of each school, in contrast to a one policy fits all approach. While resources and instructional programs are important, so is their effective implementation. Because the variance interpretation of the regression statistics more accurately represents the realities of school operations, it is the basis of estimating effect size and simulating organizational behavior.

Estimating Effect Size: Illustrations from the Minnesota Data Set

Data from Minnesota were used to examine the methods and results of measuring effect size. These results were compared with estimates from the studies reviewed in the preceding article , "A Practical Method of Policy Analysis by Considering Productivity-Related Research." This section is divided into 13 subsections.

(1) The data set
(2) Simple regression coefficients: the correlation matrix
(3) Partial correlations
(4) Method of analysis: an analytical template
(5) Regression results for teacher/pupil ratio controlled for SES
(6) Comparison with estimates from other studies
(7) Staff qualifications as an intervening variable
(8) Estimating effect size based on "value-added"
(9) Testing the Glass and Smith proposition
(10) Effect size for other staffing categories
(11) Effect size for Minnesota teacher qualifications
(12) Effect size for instructional policy options
(13) Effect size for organizational effectiveness

http://newprairiepress.org/edconsiderations/vol39/iss1/5
DOI: 10.4148/0146-9282.1102

Table 3
Correlation Matrix

	Math3	*Math5*	*Read3*	*Read5*	*SES*	*Teacher*	*Admin*	*Support*	*Aides*
Math3	1.0000								
Math5	0.7164	1.000							
Read3	0.8693	0.7568	1.0000						
Read5	0.7044	0.9286	0.7929	1.000					
SES	0.6727	0.7574	0.7609	0.8072	1.000				
Teacher	-0.3279	-0.3994	-0.3974	-0.4138	-0.5693	1.000			
Admin	-0.0033	-0.0297	-0.0079	-0.0122	-0.0011	0.0697	1.000		
Support	-0.3256	-0.3245	-0.3288	-0.3394	-0.4025	0.3467	-0.1180	1.0000	
Aides	-0.0312	-0.1197	-0.0708	-0.1030	-0.1307	0.2644	0.1126	0.0148	1.0000

The Data Set

There were some basic problems in estimating effect sizes from the Minnesota data and probably the data from most states. While the achievement scores are by grade level, the number of students and teachers are by school so that individual class sizes cannot be calculated. All other measures are also by school rather than classroom.

The data set in this analysis was constructed for another research project and is described in detail in Phelps (2009). Here I provide a summary. The data set includes 694 elementary schools over a four year period. Achievement is measured for reading and mathematics in the 3rd and 5th grades. There are data related to staffing categories and teacher qualifications. For staffing categories, these include the number of teachers, teacher aides, instructional support personnel, and administrators. Data for teacher qualifications include years of experience, salary, age, and percentage of teachers with Masters degrees. The measure of SES is in the form of an index comprised of five variables as described in Phelps (2009).

Simple Regression Coefficients: The Correlation Matrix

Table 3 presents a correlation matrix produced from the Minnesota data set. The achievement variables are: mathematics scores in 3rd grade (Math3) and 5th grade (Math5); and reading scores in 3rd grade (Read3) and 5th grade (Read5). The data for the staffing categories are measured as the staff/pupil ratio. The observations are:

- Achievement scores are highly correlated by grade and subject.
- SES is highly correlated with achievement.
- All staffing categories are negatively correlated with achievement (higher staff/pupil ratios are associated with lower achievement).
- The staffing categories are positive correlated.
- The high correlation among the staffing category variables (collinearity) poses some complexity in estimating their unique influence on achievement.

Table 4
Partial Correlations

	Math3	*Math5*	*Read3*	*Read5*
Teacher/Pupil Ratio	0.0905	0.0592	0.0671	0.0943
SES	0.5760	0.7016	0.6980	0.7269

Partial Correlations

The partial correlations for the achievement variables tell a different story. When the effect of SES is nullified (partialed out), the correlation between achievement variables and teacher/pupil ratio becomes positive. Table 4 presents the partial correlations, and the “break point,” the SES correlation coefficient where the partial correlation of the teacher/pupil ratio is zero. As the SES correlation increases, so does the partial correlation, in this case from a negative sign to a positive sign. Including some measure of SES is critical to any estimate of the influence of class size.

Method of Analysis: An Analytical Template

My original plan was to use a statistical package to run a series of regressions and report the results. This became cumbersome. While there is a great deal of information provided by statistical packages, some is devoted to making probability inferences, and the specific information needed for the policy analysis had to be moved to another setting, in this case a spreadsheet. It was possible to do the statistical calculations for the policy analysis within the spreadsheet itself. A template was created, and only the essential data required for the specific analysis was entered. Consequently, with a correlation matrix, means, and standard deviations for the essential variables, the calculations were processed and presented together in a single spreadsheet format.

Figure 2
Analysis Template to Estimate Effect Size

A	B	C	D	E	F	G	H	I
1		Correlation Matrix			Partial Correlation			
2	Variables	Math5	SES	T/P Ratio				
3	Math5	1.00000			(14.27)			
4	SES	0.75740	1.00000		0.7032			
5	T/P Ratio	-0.39937	-0.56932	1.00000	0.0593			
6								
7	Mean	1484.00	956.00	67.97				
8	Std Dev	87.73	232.00	13.28				
9								
10	Simple Regression: Y = bX + a							
11		Slope (standard partial, Beta)			R Square			
12	(15.11)	SES	0.7574		SES	0.5736		
13		T/P Ratio	-0.3994		T/P Ratio	0.1595		
14		Slope (partial, b)		Intercept (15.7)		Verification: Mean of X must produce Mean of Y)		
15	(15.5)	SES	0.2864	1210.19		1484		
16		T/P Ratio	-2.6383	1663.33		1484		
17	Multiple Regression: Y = b1X1 + b2X2 + a						Common Variance Attributed to:	
18		Slope (standard partial, Beta)		R Square		Calculation	T/P Ratio	SES
19	(16.3)	SES	0.7842	(16.5)	SES	0.59396	0.55634	0.57365
20		T/P Ratio	0.0471	(16.5)	T/P Ratio	0.01881	0.01881	0.00150
21				(16.1)	Total	0.57515	0.57515	0.57515
22		Slope (partial, b)		Intercept (16.4)				
23	(16.2)	SES	0.2965	1179.35		1484		
24		T/P Ratio	0.3111					
25		SEE (Standard error of estimate)						
26	(16.6)	57.1826						

Note: T/P Ratio = Teacher/Pupil Ratio. Std Dev = Standard Deviation.

The analytical template concentrated on the essential calculations for the later policy analysis. The policy model assumed a relationship between the policy option, in this case class size and achievement; therefore, inferential statistics were not critical. What was essential was the estimate of the magnitude of the relationship between achievement and class size, or effect size. Once the template was constructed, it was tested against a standard regression program to assure accuracy. The template consisted of two main parts: (1) Data entry comprised of the correlation coefficients, means, and standard deviations; and (2) calculations producing the regression coefficients, i.e., the weights, or effect sizes.

Statistics were calculated for simple regression (one explanatory variable) and multiple regression, with SES and teacher/pupil ratio as the explanatory variables. Simple regression results begin at B10 on the spread sheet in Figure 2, and multiple regression results begin at B17. Statistics include partial correlation coefficients; standard partial coefficients, or Beta weights; partial coefficients, or "b" weights with intercepts; the R^2, the proportion of explained variance; and standard error of estimate. Several estimates of the R^2 were provided. Verification of the functions is also included. (See G14 on the spreadsheet.) The numbers in parentheses refer to the formulae provided in Appendix A.

Regression Results for Teacher/Pupil Ratio Controlled for SES

The estimated magnitude of the relationships between the four achievement measures (mathematics and reading in the 3rd and 5th grades) and teacher/pupil ratio are presented in Table 5. The effect size estimates are the standard regression coefficients or Beta weights; b-weights with intercept; and R^2, the coefficient of multiple determination. The means of the achievement variables are also provided. From Table 5, the following observations are made:

Table 5
Effect Size Estimates for Teacher/Pupil Ratio

Coefficients	*Read3*	*Read5*	*Math3*	*Math5*	*Mean*
Teacher/Pupil Ratio					
Beta	0.0529	0.0677	0.0815	0.0471	0.0623
R Square	0.0210	0.0280	0.0267	0.0188	0.0236
SES					
Beta	0.7909	0.8457	0.7191	0.7842	0.7850
R Square	0.5597	0.6267	0.4303	0.5940	0.5527
Intercept	1198.25	1176.07	1179.35	1178.62	1183.07
SES	0.2712	0.3425	0.2965	0.2846	0.2987
Teacher/Pupil Ratio	0.3167	0.4789	0.3111	0.5635	0.4176

- SES is by far the most influential variable, explaining over half the variance, 55.27% on average, consistent with many other studies.
- When the teacher/pupil ratio is controlled for SES, the coefficient sign shifts from negative, from the correlation matrix, to positive.
- The higher the correlation between SES and achievement, the larger the teacher/pupil ratio coefficient.
- While positive, the magnitude of the relationship is small, 2.36% of the variance.

DOI: 10.4148/0146-9282.1102

Table 6
R^2 Range by Achievement Results: Common Variance Attributed to Teacher/Pupil Ratio

Is common variance attributable to teacher/pupil ratio?	*Read3*	*Read5*	*Math3*	*Math5*	*Mean*
Yes	0.0210	0.0280	0.0267	0.0188	0.0236
No	0.0019	0.0031	0.0045	0.0015	0.0027

Table 7
R^2 Estimates for Teacher/Pupil Ratio and Achievement from Hedges, Laine, and Greenwald (1994)

Beta	0.0176	0.0210	0.0176	0.0114
Estimated r	0.4	0.4	0.4	0.4
Estimated R^2	0.0070	0.0084	0.0070	0.0046

Variance is divided into two parts, the part unique to each variable and the part in common among variables. Therefore, the amount of explained variance depends on whether the common variance is attributed to SES, as is the case in regression,[9] or to teacher/pupil ratio. Table 6 presents the range when the common variance is and is not attributed to teacher/pupil ratio.

The policy implications of these results are clear: Adding teachers has a small effect on achievement. Moreover, the size of the effect depends on the inclusion of an SES variable, the weight of the SES variable, and the attribution of common variance.

Comparison with Estimates from other Studies

Hedges, Laine, and Greenwald provided estimates of the standardized regression coefficients (Betas) for teacher/pupil ratio and four estimates of effect size. These estimates have been converted to R^2 in Table 7 in order to compare them with the Minnesota estimates. The R^2 is calculated from the Beta-weight by multiplying it by the correlation coefficient between achievement and teacher/pupil ratio. The actual correlation is unknown, so a "guess-estimate" of .40 was selected.[10] These estimates are about midway between the high and low estimates from the Minnesota data.

Walberg and the Tennessee STAR experiment (Achilles 1993) provided effect size estimates. These estimates present additional problems because they are effect differences between control and experimental groups rather than standard regression coefficients. Walberg estimated the effect difference at .09 and STAR at about .24. Because there is no measure of the change in the teacher/pupil ratio, a standardized coefficient cannot be calculated directly, but an estimate can be made indirectly. (Beta is a one standard deviation change of achievement for a one standard deviation change in effect.) Assuming a one standard deviation change in the teacher/pupil ratio, the standard regression coefficients (Beta) would be .09

Table 8
R^2 Estimates from Walberg (1984) and Tennessee STAR Experiment

Studies	*Number of Standard Deviations*	*Difference*	*Correlation Coefficient*	*R Square*
Walberg	1	0.09	0.40	0.036
STAR	1	0.24	0.40	0.096
STAR	2	0.12	0.40	0.048

and .24 respectively; assuming a 2 standard deviation change for the STAR project, the Beta would be .12. Assuming a correlation coefficient with achievement of .40, the R^2 is substantially higher than the other estimates.

The Walberg estimate is about double that of the Minnesota estimate and five times higher than the analysis of Hedges et al. The Tennessee STAR estimates are substantially higher than the other two, although the 2 standard deviations assumption puts the estimates in the "ball park." These estimates will be used in the policy analysis to follow.

Staff Qualifications as an Intervening Variable

It might be possible for intervening variables other than SES to have an influence on the estimated magnitude of the class size and student achievement relationship. Data were available to test a teacher qualifications variable. Using the variables average years experience, average salary, average age, and percent of teachers with Masters degrees, a qualifications index was developed to predict mathematics achievement. Regression coefficients were applied to the data from each school to form a single index number representing the influence of these qualifications variables on achievement. The relationship between achievement and teacher/pupil ratio was calculated, including this index, with no change of results; that is, adding a qualifications index to the SES index did not improve the estimate in effect size. Because of the null results, the specifics are not reported here. Once again, the same underlying issue emerged: All variables, including variables related to teacher qualifications, are intercorrelated. Once one of the variables is included in the regression equation, it consumes the common variance and leaves little remaining unique variance for the subsequent variables.

Estimating Effect Size Based on "Value-Added"

Hanushek (2007) advocated a value-added method of production function analysis whereby value-added is achieved by inserting prior years achievement as a lag variable into the regression equation. With regard to the use of a lag variable, he stated: "Clearly, simply estimating relationships between the current level of achievement and the current inputs has little chance of accurately separating the various influences on achievement. Almost certainly, current inputs are correlated with past inputs, leading to obvious problems. The now standard approach on analyzing the growth in student achievement [the lag variable]... substantially reduces the problem" (p.168).

However, there is another consequence. Assuming that the factors influencing achievement are SES, staffing quantity, staffing qualification, and instructional materials (Phelps 2009), these factors

Published by New Prairie Press, 2017

Table 9
Effect Size Estimate for Categories of Staff-to-Pupil Ratios

Staff-to-Pupil Ratios	Math5		
	r	Beta	R^2
Teacher	-0.3994	0.0470	-0.0188
Administrator	-0.3478	-0.0289	0.0009
Support	-0.3245	-0.0234	0.0076
Aide	-0.1197	-0.0211	0.0025
SES	0.7574	0.5940	0.7842

Table 10
Estimated Range of R^2 for Minnesota Teacher Qualifications

Qualifications (expressed as averages)	R^2 Range	
	Low	High
Years of Experience	0.0073	0.0230
Salary	0.0003	0.0007
Age	0.0035	-0.0074
Percent with Masters Degree	0.0000	0.0001

Table 11
Estimated Range of R^2 for Teacher Qualifications from Hedges, Laine, and Greenwald (1994)

Qualifications (expressed as averages)	Beta		Correlation	R^2	
	Low	High		Low	High
Years of Experience	0.0414	0.0550	0.2625	0.0109	0.0144
Salary	0.0366	0.0390	-0.0445	-0.0016	-0.0017
Age	-0.0300	-0.0200	0.1102	-0.0033	-0.0022
Percent with Masters Degree	-0.0300	-0.0200	0.1102	-0.0033	-0.0022

will be present in the lag variable as well as the variables in the last time period. It is easily demonstrated that what is being measured is the difference in factors. Nevertheless, I entered a lag variable into to the regression equations for reading and mathematics at the 5th grade with little additional explanatory power, .0009 for reading and .0147 for mathematics. Because, this value-added method did not add to the measurement of effect size, it was dropped from further consideration in this analysis.

Testing the Glass and Smith Proposition: Does Achievement Improve at an Increasing Rate of Return under a Class Size of 15?

The Minnesota data have schools with class sizes lower than 15, so the Glass and Smith proposition was tested. As class sizes progressed lower than 15, predicted achievement, adjusted for SES, did not increase; in fact, it decreased slightly. It will not be considered further.

Effect Size for Categories of Staff-to-Pupil Ratios

When analyzing categories of staff-to-pupil ratios, such as those for administrators, teacher-support, and teacher-aides, the conclusions are substantially the same as for teachers. The comparison for each of the achievement measures for the four years of data were analyzed in Phelps (2009). Because the results were similar, only the data for one achievement measure, 5th grade mathematics, for one year, is presented here. (See Table 9.) In summary, for staff-to-pupil categories:

- SES explains virtually all the variance.
- The coefficient (Beta) is positive for teachers but negative for all others.
- The additional R^2 for the staffing categories is small, most likely zero for all categories except teachers.

Effect Size for Minnesota Teacher Qualifications

Minnesota data were available for the following categories of teacher qualification: Average years experience; average salary; average age; and average percentage of teachers with Masters degrees. Table 10 presents the R^2 range for these categories.

Using the method described earlier (R^2 = Beta * r), Table 11 presents the estimated R^2 for teacher qualifications from Hedges et al. The Minnesota correlations are used to calculate the R^2 from the Betas. There is a change of sign for salary because of the negative correlation.

Effect Size for Instructional Policy Options

Walberg listed estimated effect sizes for instruction, home influences, and time policies. The effect sizes are actually "effect differences" between a control group and an experimental group, and when added together, they total over 12 standard deviations. Does this mean that if all of the items were implemented by a school at the very bottom of the population (-6 standard deviations), they would progress to the very top (+6 standard deviations)? Surely not! There must be a more practical interpretation. Because of the large number of items, their conceptual similarity, and their likely intercorrelations (shared variance), they are first combined into the categories of curriculum, instructional methodology, instructional

http://newprairiepress.org/edconsiderations/vol39/iss1/5
DOI: 10.4148/0146-9282.1102

Table 12
Effect Differences and Estimated R^2 for Instructional Categories from Walberg (1984)

	Curriculum	Method	Organization	Home
Average	0.355	0.624	0.113	0.523
Beta	0.118	0.208	0.038	0.174
R^2 (r = 0.5)	0.059	0.104	0.019	0.087
Total R^2	0.269			

organization, and home influences. The average of the effect differences was calculated, reducing the standard deviation range. Second, as a matter of conjecture, two assumptions were made: The treatment difference between the control and experimental group was 3 standard deviations, so the standard regression coefficient (1 Beta) would be one-third the averaged value; and the correlation coefficient with achievement was .5 (R^2 = r * Beta). Based on these assumptions, the revised effect sizes for the categories are listed in Table 12.

With these assumptions, the R^2 are in the range of about .02 to .10, and total to approximately .27. Is there a way to determine if these estimates, or any of the other estimates, are reasonable? The next subsection provides a possible answer.

Effect Size for Organizational Effectiveness

Levin (1997) described the operations of an Accelerated School Program and presented the achievement results.[11] The overall emphasis of the program is on greater organizational effectiveness with the existing resources. For an increase of 1% in expenditures, mathematics achievement increased 45%. The information necessary to calculate an estimated effect size was unavailable although Levin claimed the influence was substantial. He identified two structural elements for consideration in a policy analysis: Incentives linked to successful performance and use of productive technology.

Building on Levin's approach, Phelps (2009) measured the potential effect size attributable to organization effectiveness. From the Minnesota data set, indices were constructed for SES, staff qualifications, staff quantity, and instructional materials. These were entered into the regression equations for the four achievement variables for each of the four years. The residuals were averaged over the four years for each observation to form a new variable, and this variable was entered into the regression equations. This process is a variation of fixed effects estimation in econometrics.[12] Schools consistently either overperformed or underperformed with regard to predicted achievement. The degree by which they missed their target is considered the measure of effectiveness.[13] The analysis also separated district effectiveness from school effectiveness. Because the analysis was of the residual and not actual data, there is no attribution to specific organizational behaviors. See Table 13 for the effect size estimates.

These estimates are valuable for several reasons:

- The measure of effectiveness--averaging of the residuals over time--substantially reduces the error variance of the equations to 0.075.
- The estimates provide an empirical base for the boundaries of effect size for the various categories of policy options described above. First, the resource-oriented variables such as staffing quantity (class size), staff qualifications (built into the salary schedules), and instructional materials seem to be limited in their overall contribution to around the average of .063. Second, the instructional and organizational variables as suggested by Walberg and Levin, do not appear to exceed the effectiveness total of .285. (The "guess-estimate" made earlier was .269.)
- The data suggest differences in the contribution of the resources and effectiveness variables based on subject matter; resources could be more important for reading, while effectiveness more important for mathematics.
- Effectiveness appears to be a shared responsibility between school and district policies and operations. This seems to imply that skilled district staff might be helpful in providing individual schools with instructional and management assistance. Moreover, good district policies would seem to support good policies in schools.

Summary and Conclusion

In this article, several achievement production models were identified stressing the importance of a unified and comprehensive operating structure, and quantifiable relationships among the elements of the structure. The studies reviewed here do not typify either a comprehensive structure or consistent measure of effect size. Based on the previous evidence and arguments presented, a fresh model emerges which provides a unifying structure, a consistent method of estimating effect size, and a coherent set of assumptions. This model emphasizes an effect ceiling and organizational effectiveness.

Table 13
Effect Size for School and District Effectiveness

Student Achievement	Without Residual	SES	Indices	District Effectiveness	School Effectiveness	Total	Error
Mathematics	0.585	0.550	0.035	0.185	0.155	0.340	0.075
Readings	0.710	0.620	0.090	0.120	0.110	0.230	0.060
Mean	0.648	0.585	0.063	0.153	0.133	0.285	0.068

The effect ceiling requires a different way of measuring effect size, while the inclusion of effectiveness variables substantially increases the accuracy of prediction. Most importantly, the model brings a new policy focus to the dilemma of Hedges, Laine, Greenwald: Why focus the primary attention on merely increasing resources (expenditures or reducing class size) if substantial achievement benefits can be derived from better instructional and organizational policies?

A Unified Structure

The reviewed research in this article focused mostly on small components of the educational process rather than treating the components as elements of a comprehensive unified structure. Class size is the primary center of attention while staffing categories other than teachers are largely ignored, counter to the notion of a team of people working together. The individual components of teacher qualifications also are viewed separately, instead of working together. Individual components of the instructional program, such as curriculum, methods, time, and instructional materials, are also viewed separately. In every case, the components are not unique or isolated; instead they are conceptually, operationally, and statistically related. An enhanced understanding of educational organizations comes from a paradigm encompassing a comprehensive system rather then reductionism to individual components.

Viewing education as a comprehensive system has implications for policy analysis. By identifying the larger categories of education and having estimates of their contribution, as well as the contribution of the component elements, it is possible to model the operation of the entire system. By simulating changes in multiple policies, the model estimates change in multiple achievement outcomes.

A unified educational structure, with its quantifiable component elements, is described in Table 14. This paradigm allows for expansion and modification of the structure to fit any circumstance where effect size and incremental cost of the policy options can be estimated. The structure that will be used in the simulation model described in the next article, "A Practical Method of Policy Analysis by Simulating Policy Options," is:

Achievement = SES+Staff Quantity+Staff Qualifications+ Instructional Program+Organizational Effectiveness

Estimating School-Specific Effect Size

The major consequence associated with the variance measure of effect size is its school-specific nature. Because the variance measure of effect size is a curve, every school will have a unique position on the curve; that is, every school will have a different marginal effect size depending on its unique circumstance. Estimating the potential of the policy options is based on seven major principles. Each principle has a different role in determining the most cost-effective policy options for the school.

Principle 1: Role of effect size. Good policy decisions start with good strategies. What is to be accomplished? How is it to be accomplished? Who is responsible? What training and mentoring is required? How will the performance and progress be monitored? Reducing class size or adding staff without first addressing these questions is foolhardy. In essence, merely adding staff without clear and comprehensive instructional (Walberg) and organizational (Levin, Phelps) strategies is counterproductive.

Principle 2: Accommodating uncertain effect size. The measurement of effect size is not precise, and research provides little in the way of reliable measures.[14] However, not all is lost. Ranges of effect sizes can be used to separate weak policy options from those with stronger possibilities. If there is a good strategy in place, then it is reasonable to assume the maximum effect size could be realized. Without a strategy, the minimum effect size is a more reasonable assumption.

Principle 3: Role of distribution variance. If effect sizes of two policy options are virtually equal, the policy with the largest variance will have the greater potential. The ability to predict is proportional to the variance; variables with larger variance are better predictors than variables with smaller variance. Other things being equal, weight should be given to the policy with the larger variance.

Principle 4: Role of the school's current status. An underlying assumption of this conceptualization is the notion of a ceiling effect—after a point, benefits for the policy option diminish. The "benefit curve" is an S-shaped curve with achievement on the Y-axis and the policy variable on the X-axis. If a school's position is low on the policy variable, the potential for improved achievement gradually increases. In contrast, if the school's position is high on the policy variable, the potential for improvement gradually diminishes.

Principle 5: Nonincremental policy options. Some policies are binary, not distributional. For example, if a new mathematics or science curriculum is based on a textbook, the policy is binary—either the textbook is adopted or it is not. Therefore principle 4 does not apply and a different method is required, which will be discussed in the next article.

Table 14
Quantifiable Component Elements of a Unified Educational Structure

Student Achievement	*SES*	*Staff Quantity*	*Staff*	*Instruction*	*Effectiveness*
Early Grades Reading Mathematics	Unique to each state	Teachers Support Aides Administration	Qualifications Education Experience Salary	Curriculum Methodology Organization Homework Time Technology	School District

DOI: 10.4148/0146-9282.1102

Principle 6: Estimating the marginal cost-effectiveness. There are three necessary numbers required to calculate the marginal cost-effectiveness of any policy option: the estimated effect size; the incremental cost; and the Z-score on the policy variable.[15] The calculation is: Effect-Size times School-Position times Marginal-Cost times.

Principle 7: Role of cost-effectiveness. If the effect sizes of two options are virtually equal, the policy with the least cost is the most cost-effective. In a complicated situation such as schools, these hand-calculations would be virtually impossible. However with current computer software, these calculations are made within fractions of a second.

References

Achilles, C.M., B.A. Nye, J.B. Zaharias, and B.D. Fulton. "The Lasting Benefits Study (LBS) in Grades 4 and 5 (1990–1991): A Legacy from Tennessee's Four-year (K–3) Class-size Study (1985–1989)." Project STAR. Paper presented at the North Carolina Association for Research in Education, Greensboro, North Carolina, January 14, 1993.

Addonizio, Michael F., and James L. Phelps. "Class Size and Student Performance: A Framework for Policy Analysis." *Journal of Education Finance* 26 (Fall 2000): 135-156.

Bohrenstedt, George W., and, Brian M. Stecher, eds. *Class Size Reduction in California: Early Evaluation Findings, 1996-98*. CSR Research Consortium. Sacramento, CA: California Department of Education, June 1999. http://www.classize.org/techreport/index.htm.

____________. *Capstone Report: What We Have Learned about Class Size Reduction in California*. CSR Research Consortium. Sacramento, CA: California Department of Education, August 2002. http://www.classize.org/techreport/index-02.htm.

Forrester, Jay W. "Systems Dynamics: Future Opportunities." In *Studies in Management Science*, edited by Augusto A. Legasto, Jay W. Forrester, and James M. Lyneis, 7-21. Vol. 14. Amsterdam, North Holland: System Dynamics, 1980.

Glass, Gene V., and Mary Lee Smith. *Meta-analysis of Research on the Relationship of Class-size and Achievement*. San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978.

Guilford, Joy Paul. *Fundamental Statistics in Psychology and Education*, 4th ed. New York: McGraw-Hill, 1965.

Hanushek, Eric A. "Some U.S. Evidence on How the Distribution of Educational Outcomes Can Be Changed." In *Schools and the Equal Opportunity Problem*, edited by Ludger Woessmann and Paul Peterson, 159-190. Cambridge, MA: MIT Press, 2007.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23 (April 1994): 5-14.

Krueger, Alan B. "Understanding the Magnitude and Effect of Class Size on Student Achievement." In *The Class Size Debate*, edited by Lawrence Mishel and Richard Rothstein, 7-35. Washington DC: Economic Policy Institute, 2002. http://edpro.stanford.edu/hanushek/admin/pages/files/uploads/classsizedebate.full%20volume.pdf.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press, 1970.

Levin, Henry M. "Raising School Productivity: An X-Efficiency Approach." *Economics of Education Review* 16 (June 1997): 303-311.

Mosteller, Frederick. "The Tennessee Study of Class Size in the Early Grades." *Future of Children* 5 (Summer/Fall 1995): 113-127. https://www.princeton.edu/futureofchildren/publications/docs/05_02_08.pdf.

Phelps, James L., "Optimizing Educational Resources: A Paradigm for the Pursuit of Educational Productivity," *Educational Considerations* 35 (Spring 2008): 3-18.

____________. "Measuring and Reporting School and District Effectiveness." *Educational Considerations* 36 (Spring 2009): 40-52.

Schrage, Linus E. *Lindo: An Optimization Modeling System*. San Francisco: CA: Scientific Press, 1991.

Walberg, Herbert J. "Improving the Productivity of America's Schools." *Educational Leadership* 41 (May 1984): 19-27.

Williams, H. Paul. *Model Building in Mathematical Programming*, 2nd ed. New York: Wiley and Sons, 1985.

Wooldridge, Jeffrey M. *Introductory Econometrics: A Modern Approach*. Cincinnati, OH: South-Western College Publishing, 2000.

Endnotes

[1] Schools have no control over the number of students, only the number of teachers.

[2] All subsequent references to Walberg in this article refer to Herbert J. Walberg, "Improving the Productivity of America's Schools," *Educational Leadership* 41 (May 1984): 19-27.

[3] The lack of a meaningful measure of SES may explain why the results from studies regarding teacher/pupil ratios and achievement are so diverse.

[4] Glass and Smith (1978) assumed an increasing return to scale and used a squared term to achieve that result. The model produced a curve with an increasing and decreasing return to scale, so they made an adjustment transforming the decreasing return to a consistent return to scale.

[5] See Phelps (2008). See also, section 3, Appendix A of this article.

[6] See the comments in the preceding article, "A Practical Method of Policy Analysis by Considering Productivity-Related Research," and Phelps (2009). This is called a fixed effect in econometrics. See also, Wooldridge (2000).

[7] All subsequent references to Hedges et al. in this article refer to Larry V. Hedges, Richard D. Laine, and Rob Greenwald, "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher* 23 (April 1994): 5-14.

[8] All subsequent references to Glass and Smith in this article refer to Gene V. Glass and Mary Lee Smith, *Meta-analysis of Research on the Relationship of Class-size and Achievement* (San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978).

[9] The variable with the highest correlation consumes the common variance.

[10] A correlation of .4 is similar to the Minnesota data, although the sign was negative in the Minnesota case.

[11] All subsequent references to Levin in this article refer to Henry M. Levin, "Raising School Productivity: An X-Efficiency Approach," *Economics of Education Review* 16 (June 1997): 303-311.

[12] See Wooldridge (2000).

[13] It is analogous to rolling a die: Some schools consistently rolled 1, 2, and 3, while others rolled 4, 5, and 6, with the target of 3.5, the average.

[14] According to Schrage (1991, 8), "The first rule of modeling is don't waste time accurately estimating a parameter if a modest error in the parameter has little effect on the recommended decision."

[15] The Z-score determines where the school is positioned on the S-shaped curve.

[16] The source for these formulae is Joy Paul Guilford, *Fundamental Statistics in Psychology and Education*, 4th ed. (New York: McGraw-Hill, 1965). Related page numbers are in parentheses.

[17] Note that the value of the correlation coefficient with the same subscript numbers, e.g., r22, is 1.

http://newprairiepress.org/edconsiderations/vol39/iss1/5
DOI: 10.4148/0146-9282.1102

Appendix A

1. Formulae for estimating effect size

Following are the formulae used to calculate the statistics in the template.[16]

1.1 Partial Correlation (14.27, p. 339):

$r_{12} = r_{12} - r_{13}\ r_{23} / \sqrt{}\ (1- r^2_{12})(1-r^2_{23})$

1.2 Coefficient for linear regression (15.55, p. 367):

$b_{yx} = r_{yx}\ (\sigma_y / \sigma_x)$

1.3 The "a" coefficient in a linear regression equation (15.7, p. 368):

$a = M_y - (M_x)b_{yx}$

1.4 Relation of regression coefficients to r2 (15.9, p. 368):

$b_{yx}\ b_{xy} = r^2$

1.5 Regression equation with standard measures (15.11, p. 370):

$Z_y = r_{yx}\ Z_x$

1.6 Standard error of estimate (15.16, p. 373):

$\sigma_{yx} = \sigma_y \sqrt{}\ (1- r^2)$

1.7 Square of coefficient of multiple correlation with three variables: (16.1, p. 394).

$R^2 = r^2_{12} + r^2_{13} - 2r_{12}\ r_{13}\ r_{23}\ / 1 - r^2_{23}$

1.8 Partial regression coefficients, the "b" weight (16.2, p. 396):

$b = (\sigma_1/\sigma_2)\ \beta_{12}$

1.9 Standard partial regression coefficients (16.3, p. 396):

$\beta_{12} = r_{12} - r_{13}\ r_{23} / 1 - r^2{}_2 3$

1.10 The "a" coefficient for linear regression (16.4, p. 397):

$a = M1 - b_{12}M_2 - b_{13}M_3$

1.11 Calculating the multiple R from Beta coefficients (16.5, p. 39):

$R^2 = \beta_{12}r_{12} + \beta_{13}r_{13}$

Note that if the correlation is negative, the absolute value is taken. However, the result is not consistent with equation 16.1. Actually the R^2—the proportion of explained variance—is divided into two parts, the unique part and a common part. Equation 16.5 attributes both the unique and common parts to each variable, thus the sum is larger than 16.1. As a result, a choice must be made as to which variable will receive the common variance. The unique variance of the remaining variable is calculated by subtracting the unique and common variance of the selected variable from the R^2 from equation 16.1:

$R^2 - \beta_{12}R_{12} = \beta_{13}R_{13}$

This is consistent with the principles of stepwise regression. The first term in (with the highest correlation with the outcome variable) assume both the unique and common variance with the other variables. The next variable in assumes just the unique variance.

1.12 Standard error of multiple estimate (16.6, p. 400):

$\sigma_{yx} = \sigma_y \sqrt{}\ (1- R^2)$

1.13 Multiple regression with more than three variables (16.13, p. 409)

Each time a variable is added to the regression equation, the Betas must be recalculated. The calculation answers the question: What regression weights would best predict the outcome variable from the explanatory variables? The calculation is based on normal equations, with one fewer equation than the number of variables in the equation (including the outcome). The solution to these normal equations can be found by employing a software program, like Microsoft Excel's Solver. The follow example can be expanded to include any number of variables.

$r_{22}\ \beta_{12} + r_{32}\ \beta_{13} + r_{42}\ \beta_{14} = r_{12}$

$r_{23}\ \beta_{12} + r_{33}\ \beta_{13} + r_{43}\ \beta_{14} = r_{13}$

$r_{24}\ \beta_{12} + r_{34}\ \beta_{13} + r_{44}\ \beta_{14} = r_{14}$

2. Converting standard regression coefficients to R^2

The following principles apply. If a value is unknown, then an estimate must be made to stay within the principles.

2.1. The total of all the variance is 1: $R^2 = 1$

2.2. The R^2 for the individual explanatory variables is calculated by the formula:

$R^2 = \beta_{12}r_{12} + \beta_{13}r_{13} + \dots\ \beta_n r_n$

Appendix A (continued)

2.3. The estimated range of the nonresource explanatory variables is:

SES = 55 to 60; Error 7 to 10

Effectiveness (instructional and organizational) 25 to 27.

2.4. The range for the resources explanatory variables, therefore, must be between 3 and 13.

3. Interpretation of Variance

Statistical variance is a general term referring to the area under the normal distribution, but it is measured in two ways. The first method is in terms of square units, and the second is in terms of a linear parameter of the normal distribution. It is important to distinguish between the two measures because the same word, variance, is used to describe both concepts. The focus here is on how variance can be the bases of estimating effect size.

3.1. The sum of squared deviations from the mean of the distribution gives a measure of the total area under the distribution, or total variance area.

3.2. The parameter of the distribution is calculated by taking the average squared deviation, also called the variance, or σ^2, the square root of which is the standard deviation or σ. The standard deviation is the width parameter of the distribution. The standard deviation is also the parameter in determining the area under the normal curve: $\sigma\sqrt{2\pi}$.

3.3. The principle of regression is to find a line for which the sum of the squared deviations (area) around the line is a minimum. This is the error variance area. Because the regression line is the mean of the distribution, the standard error of estimate is the standard deviation or width parameter of the distribution around the line (p.375). In other words, the total variance area is comprised of the explanatory variance area and error variance area.

3.4. Divided equation (3.3) by the total variance area, the results are percentages, the percentage attributable to the explanatory variables and error. Because the total percentage is 1.00, the percentage of the explanatory variance area (that explained by the regression line) and error variance area are:

$$1 = \%\ \text{Explanatory Variance Area} + \%\ \text{Error Variance Area}$$

3.5. Regression programs provide these sum of the square numbers from which the explanatory variance area is calculated. It is said the explanatory variable explains a certain proportion of the total variance. It is called the coefficient of determination, and noted as the R^2.

3.6. Each explanatory variable has a unique R^2 based on the relationship between the Beta and correlation coefficient:

$$R^2 = \beta_{12}r_{12} + \beta_{13}r_{13}$$

3.7. As additional explanatory variables are added, as is the case in stepwise regression, the amount of explanatory variance increases to a maximum point.

3.8. The area of the normalized curve is 1; therefore the proportion of variance explained by each component, explanatory variables and error (or residual), sum to 1.00 with the R^2 for each component representing a percentage of area under the normal curve.

3.9. The percentage area of each component can be converted to the cumulative area under the normal curve or percentile. This curve is S-shaped with asymptotes at 0 and 100 percentiles. Because the mean of the explanatory variable equals the mean of the outcome variable, one-half of the R^2 area is above the 50th percentile and one-half below. For example, if the R^2 is .50, the asymptotes are at the 25th and 75th percentiles.

4. Calculations for the normal curve and area under the curve

4.1 The equation for the normal curve is:

$$Y = e\ \exp\text{-}Z^2/2\ /\ \sqrt{\ }\ (2\ Pi)$$

The cumulative area under the normal curve is the integral of the normal curve. Therefore, the slope of the integral at any point is calculated via the normal curve equation by inserting the value of Z.

http://newprairiepress.org/edconsiderations/vol39/iss1/5
DOI: 10.4148/0146-9282.1102

Appendix B

Summary of Effect Sizes Converted to R^2

Tables B-1 and B-2 summarize the materials presented in the body of this article. In Table B-1, the effect sizes are presented in terms of the amount of variance explained or the R^2. In some cases, a conversion was made from the original metric to the R^2 metric, based on the formulae described previously. The summary is presented in three major categories: Staffing; instruction; and qualifications. Each of the categories includes the associated elements. For each of the studies reported, a low and high estimate are presented. When the correlation or Beta coefficient is negative, the results are presented as negative.

In Table B-2, summary calculations are provided. For each category and element an average low, average high, and average are calculated. In order to evaluate the estimates, the absolute values are calculated and then totaled to determine their total explanatory value, the total of which cannot exceed 1.00, including error. The Staffing category ranged from .0437 to .0587; Instruction ranged form .1523 to .2700; and Qualifications from .0178 to .0240. The totals for these categories ranged from a low of .1870 to a high of .3527, with the average of .2640. When the R^2 of SES is set as .5800 (from the Minnesota data), the error contribution is calculated.

When these data are taken together, the ranges are similar to the results obtained from the analysis of the Minnesota data set. Importantly, these data reflect the product of a methodology to estimate a consistent effect size from studies with different measures. These are not intended to represent a definitive estimate. Nevertheless, these estimates are thought to be a reasonable starting point for use in a simulation model.

Summary Table B1
Effect Sizes from Various Studies

Variables	Minnesota		Hedges et al.		Krueger		Walberg		STAR		California CSR	
	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Staffing												
Teacher/Pupil Ratio	0.0015	0.0188	0.0070	0.0080	0.0800	0.0800	0.0360	0.0450	0.0400	0.1000	0.0000	0.0400
Support/Pupil Ratio[1]	-0.0076	0.0005										
Aide/Pupil Ratio	-0.0025	0.0004							0.0000	0.0000		
Administrator/Pupil Ratio	-0.0032	0.0001										
Instruction												
Curriculum							0.0235	0.0470				
Method							0.0415	0.0830				
Organization							0.0015	0.0030				
Homework							0.0350	0.0700				
Time							0.0255	0.0510				
Qualifications												
Experience	0.0073	0.0230	0.0109	0.0144								
Salary	0.0003	0.0007	-0.0016	-0.0017								
Masters Degree	0.0000	0.0001	-0.0033	-0.0022								
Age	0.0035	-0.0074										

[1] "Support" refers to instructional support personnel such as reading teachers.

Appendix B (continued)

Table B2
Summary Calculations

Variables	*Average*		*Average*	*Absolute Value*	*Subtotal*	*Absolute Value*		*Subtotal*
	Low	*High*				*Low*	*High*	
Staffing								
Teacher/Pupil Ratio	0.0329	0.0584	0.0380	0.0380		0.0329	0.0584	
Support/Pupil Ratio[2]	-0.0015	0.0001	-0.0036	0.0036		0.0015	0.0001	
Aide/Pupil Ratio	-0.0013	0.0002	-0.0005	0.0005		0.0013	0.0002	
Administrator/Pupil Ratio	-0.0032	0.0001	-0.0016	0.0016		0.0032	0.0001	
					0.0437			0.0587
Instruction								
Curriculum	0.0295	0.0590	0.0353	0.0353		0.0295	0.0590	
Method	0.0520	0.1040	0.0623	0.0623		0.0520	0.1040	
Organization	0.0100	0.0200	0.0023	0.0023		0.0100	0.0200	
Homework	0.0435	0.0870	0.0525	0.0525		0.0435	0.0870	
Time			0.0383	0.0383				
Qualifications								
Experience	0.0091	0.0187	0.0139	0.0139		0.0091	0.0187	
Salary	-0.0007	-0.0005	-0.0006	0.0006		0.0007	0.0005	
Masters Degree	-0.0017	-0.0010	-0.0013	0.0013		0.0017	0.0010	
Age	0.0017	-0.0037	-0.0020	0.0020		0.0017	0.0037	
					0.0178			0.0240
		Subtotal		0.2520	0.2138	0.1870	0.3527	0.3527
		SES			0.5800	0.5800	0.5800	
		Total			0.7938	0.7670	0.9327	
		Error			0.2062	0.2330	0.0673	
		Grand Total			1.0000	1.0000	1.0000	

[2] "Support" refers to instructional support personnel such as reading teachers.

http://newprairiepress.org/edconsiderations/vol39/iss1/5
DOI: 10.4148/0146-9282.1102