How Practical is Value-Added Analysis for Estimating Treatment Effects in School Settings?

FJ King and Yuwadee Wongbundhit

Florida State University

Yvonne M. Bigos

Leon County Schools

B. F. Foster

Florida Department of Education

One concern facing program evaluators is that of selecting appropriate models for estimating treatment effects. This is particularly true in school settings where it is not possible to obtain comparable control groups. In such cases it is difficult to separate the amount of growth that occurs as a result of the treatment from the amount of growth that occurs naturally over time in the absence of the treatment. A seemingly attractive method which allows an evaluator to separate these two components of growth is use of the value-added model (Bryk & Weisberg, 1976; Bryk, Strenio, & Weisberg, 1980).

The basic idea underlying this model is that the effect of an intervention can be estimated by comparing the average observed growth between pretest and posttest with the estimated growth expected in the absence of the intervention. An unbiased estimate of the average growth rate can be provided via ordinary least squares regression of pretest on age if age at the time of pretest is not related to systematic growth. When growth (estimated cross-sectionally) is not linear, transformations must be made before value-added analysis can be completed.

Bryk, et al. (1980), pointed out that violation of the assumptions of independence of growth and age in cross-sectional data will generally result in a non-linear relationship between pretest and age. Such violations can occur in at least two different ways. The first is that historical trends may cause differences among groups of children. For example, some event that occurred in the past could cause children conceived during that period to develop at different rates from other younger or older children. The second, and probably more plausible, reason for violation of the assumption is selection.

Individual school promotion and retention policies and the probabilities that precocious children enter school at earlier ages than do children with slower growth rates could cause the assumption, even though true for the total population, to be violated in cross-sectional samples.

In the model's simplest form, pretest scores are regressed on age and the regression coefficient (b) is taken to be an unbiased estimate of growth rate. The intervention effect, or value added by the treatment (V), is then

$$V = \bar{y}_2 - \bar{y}_1 - b(a_2 - a_1)$$

where $\bar{y}_1$ and $\bar{y}_2$ are the pre- and posttest means, and a_1 and a_2 are mean ages at pre- and posttest times.

Unfortunately, no test of significance of V exists. Bryke, et al. (1980), however pointed out that the jackknife technique (Mosteller & Tukey, 1977) can be used to provide a test statistic. It involves the computation of a psuedo-value V_i^* for each individual in the sample, treating these values as data points, and calculating their mean and standard errors. The mean is an unbiased estimate of V and the standard error allows the calculation of a t-test (df = N-1) for significance testing or interval estimation. The following steps are involved in the computation of the psuedo values.

1. Compute the regression coefficient (b) using the whole data set.

2. Compute regression coefficient (b_i) with observation i removed from the data. N coefficients will be computed.

3. Compute a psuedo value (V_i^*) for each individual:

$$V_i^* = y_i(t_2) - y_i(t_1) - b_i^* [a_i(t_2) - a_i(t_2)]$$

where $y_i(t_2)$ and $y_i(t_1)$ are posttest and pretest scores for individual i, $a_i(t_2)$ and $a_i(t_1)$ are the ages of individual i at posttest and pretest times, and b_i^* is computed as shown below:

$$b_i^* = Nb - (N - 1)b_i$$

4. Compute the mean and standard error of the V_i^*'s. Calculate a t-ratio by dividing the mean by the standard error.[1]

The model can be extended to incorporate background variables that may be related to individual growth rates. Computationally this is done by regressing the pretest on age and the first order interactions of age and each background variable. Examples of such variables given by Byrk, et al. (1980), include child's sex (1 = male, -1 = female), child's race (1 = black, -1 = white) and mother's education (1 = more

---

[1]A set of SPSS procedures that eliminates the need for computing N regression equations is shown in the Appendix.

than 12 years, -1 = 12 years or less). With dichotomous background variables the resulting regression equation is,

$$Y_2 = b_0 + b_1 a_1 + b_2 a_1 x_1 + b_3 a_1 x_2 + b_4 a_1 x_3$$

where Y_2 is the posttest, a_1 is age at pretest time and the x's are sex, race, and mother's education. This equation can be decomposed to yield separate, single estimates of the regression of pretest on age for every combination of background variables. The calculation of the value-added estimates can then be made although estimating the psuedo values (V_i^*) would probably require N separate regressions or the construction of a special computer program to obtain them.

The purpose of this study was to investigate the use of value-added analysis as a means of estimating treatment effects in several sets of data which varied with respect to the age level of the students involved and the outcome measures used. Of primary interest was whether non-linear relationships would be found between age and pretest scores and if so, whether transformations could be found to linearize them.

Value-added analyses were attempted on one data set from pre-school children in a compensatory education program, and four sets of data involving fifth and sixth grade students. The first step in each analysis was to test the assumption that the relationship of age to pretest achievement is linear. If a significant non-linear relationship were found, it could indicate a violation of the key assumption of the method - that is, within the sample under consideration "individual growth characteristics are independent of age." Following the finding of a significant non-linear relationship, the next step was to determine its form and to seek a way to transform it to one which was at least approximately linear. Finally, the mean value added by the treatment and a test of its significance were estimated when a linear relationship was found.

Analysis of Pre-School Data[2]

The 47 students in this study ranged in age from 27 to 56 months at the time the pretest was given. They were enrolled in two schools in pre-kindergarten programs designed to prepare educationally disadvantaged children in readiness skills. The program was implemented over a seven-month period.

The test used to evaluate the program was the Learning Accomplishment Profile (LAP) (Lemay & Maltes, 1976), an individually administered developmental test. The four separate skill areas assessed in this study were fine motor manipulation (FM), gross motor movement (GM), cognition (C) and language (L).

The first analysis tested the assumption of linearity of the regression of age on the four pretest measures. Table 1 shows the squared correlations between each measure and age and age squared. None of the differences was significant. Hence, the conclusion was that for this sample of students, the linearity assumption was met.

The linear equations for the regression of the pretest measures on age were as follows:

---

[2]These data were furnished to us by Dr. Barbara Foster, Director of Title I Evaluation, Florida State Department of Education.

Table 1

Squared Correlations between Pretest Measures and Age and Pretest Plus Age Plus Age Squared

Pretest	Squared Correlations: Age	Age + Age^2	Differences
Fine Motor	.609	.612	.003
Gross Motor	.577	.586	.010
Cognition	.707	.707	.000
Language	.717	.717	.000

$$\hat{FM} = 6.48 + .91 \text{ age}$$

$$\hat{GM} = 6.85 + 1.07 \text{ age}$$

$$\hat{C} = -27.79 + 1.56 \text{ age}$$

$$\hat{L} = 15.14 + .92 \text{ age}$$

An estimate of the value added by the treatment for each individual (V_i^*) for each skill area was calculated and the mean, standard error and t-ratio were computed for each variable. The results are shown in Table 2. Three of the skill areas give evidence of having been enhanced by the treatment. Gross motor movement shows no evidence of growth beyond natural maturation. By contrast, analysis of pretest-posttest gains showed significant results for all four skill areas.

Table 2

Mean Estimates of Value Added ($\overline{V}_i^*$), Standard Errors and T-ratios for Four Skill Area Measures

Measure	$\overline{V}_i^*$	Standard Error	t-ratio
Fine Motor	4.15	.99	4.18*
Gross Motor	1.08	1.27	.85
Cognition	10.99	2.23	4.93*
Language	4.81	1.18	4.08*

*Probability less than .01 (df = 46)

Analysis of Fifth and Sixth Grade Data Sets

Three of the fifth and sixth grade data sets were from counties in North Florida that were participating in a Title IV-C reading curriculum development program. Approximately equal numbers of the students in each set were treatment and control subjects. The pre-posttests consisted of multiple-choice cloze reading passages that encompassed a wide range of difficulty. Each group received tests based on different subject matter content. None of the students was participating in compensatory education programs but some classes contained "main-streamed" students, i.e., students who were formerly classified as educable mentally retarded but who were currently in the regular school program. Pretests were administered in the winters of 1978 and 1979; posttests were given six weeks after each pretest administration.

The fourth data set contained students who were participating in an ESAA project. They were tested with the Metropolitan 1978 reading comprehension subtest in October, 1979, and April, 1980.

Table 3 shows the results of the first step of the analysis. It contains sample sizes, regression equations, probabilities, and adjusted squared multiple correlations for the four data sets.

Table 3

Sample Sizes, Regression Equations, Probabilities and Adjusted Squared Multiple Correlations for the Fifth and Sixth Grade Data Sets

				Regression of Pretest on Age				
				Age		Age Squared		
Grade	Test	N	Constant	b_1	Prob.	b_2	Prob.	R^2
5th	Cloze 1	198	-196.65	3.69	.06	-.014	.05	.06
6th	Cloze 2	80	-105.36	2.14	.33	-.007	.32	.02
5th/ 6th	Cloze 3	316	-195.51	3.33	.004	-.011	.003	.03
5th/ 6th	Metro	221	-286.55	5.49	.09	-.023	.06	.11

The results show that all four of the equations have the same form although the probabilities for the sixth grade sample of size 80 failed to approximate those usually considered to be statistically significant. For all four data sets the regression curve was parabolic within the observed age range. That is, it specified positive growth at lower ages but negative growth at the upper level.

These results clearly indicate that the assumption of independence of growth characteristics and age is violated in these samples.

Furthermore, no transformation was found that could linearize these curves.

The possibility that selection processes were responsible for the non-linear relationship found between age and pretest led us to follow the example of Byrk, et al. (1980), to eliminate from our largest sample of fifth and sixth grades 16 students whose ages were greater than 14 years and 3 months. In addition, we eliminated 52 students whose I.Q.'s (SFTAA) were below 85, on the assumption that their growth rates were probably slower than those of most students. Eliminating students on the basis of prior classification as EMR rather than on I.Q. would have been preferable, but such data were unavailable.

The remaining 248 cases were reanalyzed and it was found that the regression of pretest on age still gave evidence of non-linearity. However, the regression curve exhibited only a slight downward turn at the upper age levels. The negative exponential function given by Bryk, et al. (1980), was used to transform the pretest data (ln [1-pre/K] where K is the asymptote value, in this case the number of items in the test plus 1, i.e., 81) and the resulting regression of pretest on age appeared to be sufficiently linear to proceed with the analysis. The regression equations for the reduced samples are given in Table 4.

The 248 students were divided into treatment and control groups, posttest scores were transformed, and the value-added effects were computed separately for each group. The jackknife procedure suggested by Byrk, et al. (1980), was used to test the significance of the value added for each group separately.

V_i^* means for the treatment and control groups were .07 and .04 with standard errors of .04 and .02 respectively. The confidence

Table 4

Regression Equations, Probabilities and Adjusted Squared Multiple Correlations for Reduced Sample

Dependent Variable	N	Constant	Regression Equation: Age b_1	Age Prob.	Age Squared b_2	Age Squared Prob.	R^2
Pretest	248	-177.06	2.94	.07	-.009	.09	.04
Transformed Pretest	248	-.135	.010	.001	--	--	.05

intervals (95%) for the treatment group were, -.01 to .14, and -.01 to .08 for the control group. Thus, no evidence for a treatment effect could be claimed. In contrast, a regression analysis of the non-equivalent control group design for both the original sample and the reduced sample with transformed data showed significant treatment and treatment by pretest interaction effects. The results indicated that students with high pretest scores profited from the treatment but that students with low pretest scores did not (standardized effect sizes at one standard deviation above and below the mean were .37 and -.19 respectively for the original data).

Discussion

The results of this study agree with those of Bryk and Weisberg (1976) and Bryk, Stevens and Weisberg (1980) in suggesting that with pre-school children, the assumptions of the model are likely to be met. However, these assumptions should be verified for each analysis. Value-added analysis should be especially useful with tests that have no national norms such as the LAP, and in cases where no control groups are available. The jackknife technique may not, at present, be widely known to local school district evaluators and its use will present problems in situations where computer facilities are not adequate. This is especially true when background variables are incorporated into the model.

Our results suggest that the use of the value-added model in studies where the subjects are older school age children may not be appropriate. In all of the fifth and sixth grade data sets, the relationship between pretest and age was weak and non-linear. This could result from selection processes, as previously indicated, or from the possibility that, at the age levels involved, longitudinal growth is not linearly related to age.

One data set was transformed into a linear relationship, but only after eliminating a substantial number of subjects from the analysis. In addition, the results of the value-added analysis disagreed with the regression analysis of posttest on pretest, treatment and their interaction, possibly because of the significant interaction effect.

None of the original evaluation plans for our fifth and sixth grade data sets called for a value-added analysis. It is possible that in locations where selection processes are different and/or where background variables that are related to individual growth can be incorporated into the model different results would be found.

References

Bryk, A. S., Strenio, J. F., & Weisberg, H. I. A method for estimating treatment effects when individuals are growing. Journal of Educational Statistics, 1980, 5(1), 5-34.

Bryk, A. S., & Weisberg, H. I. Value-added analysis: A dynamic approach to the estimation of treatment effects. Journal of Educational Statistics, 1976, 1(2), 127-155.

Hildreth, G. Results of repeated measurement of pupil achievement. Journal of Educational Psychology, 1930, 21, 286-296.

LeMay, D. W., & Maltes, S. Learning accomplishment profile technical report. Chapel Hill, N.C.: Chapel Hill Training-Outreach Project, 1976.

Mosteller, F., & Tukey, J. W. Data analysis and regression. Reading, Mass.: Addison-Wesley, 1977.

Appendix

1. Computation of psuedo values (V_i^*) for the fine motor subtest of LAP.

   X = age (months), Y1 = FM pretest (pre FM), N = 47

   Y_2 = FM posttest, b = $(N\Sigma XY - (\Sigma X)(\Sigma Y))/(N\Sigma X^2 - (\Sigma X)^2)$

2. Use SPSS condescriptive to compute the following:

   Compute; SXY = age * pre FM

   Compute; SX^2 = age * age

   Condescriptive; SXY, SX2, SX, SY

   Statistics; 12

3. The output of the condescriptive run yields the following sums:

   SXY = 108150 = ΣXY SX = 2179 = ΣX

   SX2 = 103249 = ΣX^2 SY = 2289 = ΣY

4. Use SPSS condescriptive and the sums obtained above to compute the following:

   Compute; SXY = 109150

   Compute; SX2 = 103249

   Compute; SX = 2179

   Compute; SY = 2289

   Compute; B = .9107 = coefficient computed from the whole data set

   Compute; Num = 46 * (SXY - (X * Y)) - ((SX-X) * (SY-Y))

   Compute; Denom = (46 * (SX2 - X**2)) - (SX-X) **2

   Compute; BI = Num/Denom = bi

   Compute; BSI = (47 * B) - (46 * BI) = b*i

   Compute; V = Y2 - Y1 - (BSI * 7) = V_i^*, 7 = number of months of treatment

Condescriptive; V

Statistics; 1, 2

5. The second condescriptive run yields the mean of the psuedo values ($\overline{V}_i^*$) and its standard error.