Value Added Productivity Indicators:  
A Statistical Comparison of The  
Pre-Test/Post-Test Model and Gain Model *  

Dash Weerasinghe  Timothy Crsak  Robert Mendro  

Dallas Public Schools  
Dallas, TX 75204  

March 4, 1997  

Abstract  

In an age of student accountability, public school systems must find procedures for identifying effective schools, classrooms and teachers that help students continue to learn academically. As a result researchers have been modeling schools and classrooms to calculate productivity indicators that will withstand not only statistical review but political criticisms. There are two main approaches to this problem; one using Pre-test and Post-test scores of students while the other uses student gain scores. In this paper a model is developed for each approach and their results are discussed. The outcomes of each model are School Effective Indices(SEI) and Classroom Effective Indices(CEI). A set of criteria is established and the SEI's the two models are tested against these criteria to determine the "better" approach.  

Introduction  

The strategy for predicting performance from a pre-test score with adjustments for covariates is becoming the standard for producing productivity indicators. Currently the most asked question is: should we be modeling the student’s post-test score or student growth, i.e., gain.  

To model student growth or student achievement within a year, the standard statistical approach is to model an outcome variable measuring student’s achievement or growth in a  

regression model as follows: If ITBS96_{ij} is the i^{th} student's score on ITBS Reading in school j for year 1995/96, then

\[ \text{ITBS96}_{ij} = \beta_0 + r_{ij} \]  

where \( r_{ij} \) are the student residuals which are \( N(0, \sigma) \).

In this model, \( r_{ij} \) measures the distance a student’s actual score varies from the regression line. If \( r_{ij} > 0 \), the student performed higher than average and if \( r_{ij} < 0 \), the student performed lower than average. By aggregating the \( r_{ij} \) by each \( j \), we can calculate a measure of how well school \( j \) performed. The question of interest is what does \( r_{ij} \) measure. We are interested in measuring student achievement in the year 1995/96, but in this particular model \( r_{ij} \) is not a measure of student growth, but current standing. This can be easily observed by the bias of \( r_{ij} \) with respect to student’s previous test scores.

The next progression in this model is to remove any biases in \( r_{ij} \) with respect to students’ prior test scores, namely ITBS Reading scores from school year 1994/95. Hence equation 1 is expanded as follows:

\[ \text{ITBS96}_{ij} = \beta_0 + \beta_1 \text{ITBS95}_{ij} + r_{ij} \]  

Now, \( r_{ij} \) in the model is conditioned on student’s pre-test score and can be said to be unbiased with respect to the student’s 1994/95 ITBS reading score. Now, \( r_{ij} \) is a measure of student growth in year 1994/95/96.

The next major question arises at this stage: what is the outcome variable that should be modeled? Should it be the student’s ITBS Reading score for 1995/96 or should it be the gain in the student’s ITBS Reading score from 1994/95 to 1995/96? To answer this question, two types of models will be analyzed, the one used by a large urban school district and a complex gain model recently developed (Bryk and Thum, April 1996). The Bryk and Thum gain model is as follows:

\[ \text{GAIN9596}_{ij} = \beta_0 + \beta_1 \text{ITBS95}_{ij} + r_{ij} \]  

For this study, the measure GAIN9596 will be calculated as follows:

\[ \text{GAIN9596}_{ij} = \text{ITBS96}_{ij} \cdot \text{GE}_{ij} - \text{ITBS95}_{ij} \cdot \text{GE}_{ij} \]  

In this study data from 5197 sixth grade students from 118 elementary schools were modeled, with no missing data.

**Methodology**

Hierarchical Linear Modeling will be used to compare the Pre-Test/Post-Test Model and the Gain Model. This comparison will be carried out for four models of varying complexities. The school rankings obtained from the two types of models will be compared to each other to observe any differences.

The four models are as follows:

2
The basic models with no student level variables and no school level variables:

**PP-1 Model:**

\[
\text{READ}_{96ij} = \beta_{0j} + \beta_{1j} \text{READ}_{95ij} + r_{ij} \tag{5}
\]
\[
\beta_{0j} = \gamma_{00} + u_{0j} \tag{6}
\]
\[
\beta_{1j} = \gamma_{10} \tag{7}
\]

**Gain-1 Model:**

\[
\text{READGAIN}_{9596ijk} = \pi_{0jk} + \pi_{1jk} \text{READ}_{95ijk} + e_{ijk} \tag{8}
\]
\[
\pi_{0jk} = \beta_{00k} + \gamma_{0jk} \tag{9}
\]
\[
\pi_{1jk} = \beta_{10k} + \gamma_{1jk} \tag{10}
\]
\[
\beta_{00k} = \gamma_{000} + u_{00k} \tag{11}
\]
\[
\beta_{10k} = \gamma_{100} \tag{12}
\]

The basic models with no student level variables and one school level variable, percentage of minority students in a school, \( \text{PMINORITY} \):

**PP-2 Model:**

\[
\text{READ}_{96ij} = \beta_{0j} + \beta_{1j} \text{READ}_{95ij} + r_{ij} \tag{13}
\]
\[
\beta_{0j} = \gamma_{00} + \gamma_{0j} \text{PMINORITY}_{j} + u_{0j} \tag{14}
\]
\[
\beta_{1j} = \gamma_{10} \tag{15}
\]

**Gain-2 Model:**

\[
\text{READGAIN}_{9596ijk} = \pi_{0jk} + \pi_{1jk} \text{READ}_{95ijk} + e_{ijk} \tag{16}
\]
\[
\pi_{0jk} = \beta_{00k} + \gamma_{0jk} \tag{17}
\]
\[
\pi_{1jk} = \beta_{10k} + \gamma_{1jk} \tag{18}
\]
\[
\beta_{00k} = \gamma_{000} + \gamma_{01k} \text{PMINORITY}_{k} + u_{00k} \tag{19}
\]
\[
\beta_{10k} = \gamma_{100} \tag{20}
\]
Models with student level variables (GENDER, BLACK and HISPANIC) and no school level variables:

**PP-3 Model:**

\[
\text{READ96}_{ij} = \beta_0 + \beta_1 \text{READ95}_{ij} + \beta_2 \text{SEX}_{ij} + \beta_3 \text{BLACK}_{ij} + \beta_4 \text{HISPANIC}_{ij} \\
+ \beta_5 \text{SEX}_{ij} \cdot \text{READ95}_{ij} + \beta_6 \text{BLACK}_{ij} \cdot \text{READ95}_{ij} \\
+ \beta_7 \text{HISPANIC}_{ij} \cdot \text{READ95}_{ij} + \epsilon_{ij}
\]  

(21)

\[
\beta_{ij} = \gamma_{00} + \gamma_{0q} u_{ij}
\]

(22)

\[
\beta_{qj} = \gamma_{0q}
\]

(23)

for \( q = 1, 2, \ldots, 7 \).

**Gain-3 Model:**

\[
\text{READGAIN9596}_{ijk} = \pi_{0jk} + \pi_{1jk} \text{READ95}_{ijk} + \epsilon_{ijk}
\]  

(24)

\[
\pi_{0jk} = \beta_{00} + \beta_{10} \text{SEX}_{jk} + \beta_{20} \text{BLACK}_{jk} + \beta_{30} \text{HISPANIC}_{jk} + \epsilon_{0jk}
\]

(25)

\[
\pi_{1jk} = \beta_{01} + \beta_{11} \text{SEX}_{jk} + \beta_{21} \text{BLACK}_{jk} + \beta_{31} \text{HISPANIC}_{jk} + \epsilon_{1jk}
\]

(26)

\[
\beta_{00} = \gamma_{000} + \gamma_{0q0} u_{jk}
\]

(27)

\[
\beta_{0q} = \gamma_{0q0}
\]

(28)

for \( p = 0, 1 \), \( q = 1, 2, 3 \), and \( p = q \neq 0 \).

Models with student level variables (GENDER, BLACK and HISPANIC) and one school level variable, percentage of minority students in a school, \( \text{PMINORITY} \):

**PP-4 Model:**

\[
\text{READ96}_{ij} = \beta_0 + \beta_1 \text{READ95}_{ij} + \beta_2 \text{SEX}_{ij} + \beta_3 \text{BLACK}_{ij} + \beta_4 \text{HISPANIC}_{ij} \\
+ \beta_5 \text{SEX}_{ij} \cdot \text{READ95}_{ij} + \beta_6 \text{BLACK}_{ij} \cdot \text{READ95}_{ij} \\
+ \beta_7 \text{HISPANIC}_{ij} \cdot \text{READ95}_{ij} + \epsilon_{ij}
\]  

(29)

\[
\beta_{ij} = \gamma_{00} + \gamma_{0q} \text{PMINORITY}_{j} + \epsilon_{ij}
\]

(30)

\[
\beta_{qj} = \gamma_{0q}
\]

(31)

for \( q = 1, 2, \ldots, 7 \).
Gain-4 Model:

\[ \text{READGAIN9596}_{ijk} = \pi_{0ijk} + \pi_{1jk} \text{READ95}_{ijk} + \epsilon_{ijk} \]  

\[ \pi_{0ijk} = \beta_{00k} + \beta_{01k} \text{SEX}_{jk} + \beta_{02k} \text{BLACK}_{jk} + \beta_{03k} \text{HISPANIC}_{jk} + \tau_{0jk} \]  

\[ \pi_{1jk} = \beta_{10k} + \beta_{11k} \text{SEX}_{jk} + \beta_{12k} \text{BLACK}_{jk} + \beta_{13k} \text{HISPANIC}_{jk} + \tau_{1jk} \]  

\[ \beta_{00k} = \gamma_{000} + \gamma_{001} \text{PMINORITY}_{k} + u_{00k} \]  

\[ \beta_{pqk} = \gamma_{100} \]  

for \( p = 0, 1, q = 1, 2, 3, \) and \( p = q \neq 0. \)

For models PP-1, PP-2, FP-3 and PP-4:

\( i = 1, 2, \ldots, 5197 \) and \( j = 1, 2, \ldots, 118. \)

Ranking of schools are obtained from \( \bar{u}_{ij} \) which measures the deviation of the individual school’s intercept from the overall intercept of the schools. This estimate is shrinkage adjusted by the number of repeated observations used to calculate this measure.

For models GAIN-1, GAIN-2, GAIN-3 and GAIN-4:

\( i = 1, 2, \ldots, 5197, \quad j = 1, 2, \ldots, 5197 \) and \( k = 1, 2, \ldots, 118. \)

Ranking of schools are obtained from \( \bar{u}_{ijk} \) which measures the deviation of the individual school’s intercept from the overall intercept of the schools. This estimate is also shrinkage adjusted by the number of repeated observations used to calculate this measure.

Classroom Effective Indices are obtained from the student residuals, namely \( r_{ij} \), for the PP models and \( r_{ijk} \), for the Gain Models, which measure the deviation of the students score from its predicted value.

**Results**

**School Effective Indices**

Under the model assumptions, the \( u_{ij} \)'s for the 2-level models and \( u_{ijk} \)'s for the 3-level models are normally distributed with a mean of zero. This fact is necessary if the school rankings obtained from the models are used to rank schools, and specially if the top schools are given performance awards. The need may arise to calculate confidence intervals and the normality assumptions will be vital. The descriptive statistics for the measures of School Effective Indices are as follow: 

5
Kolmogorov-Smirnov Goodness of fit test was carried out to determine if these measure are Normally distributed. The table above indicates K-S P-Values which are extremely high, hence failing to reject the Null hypothesis that the values are from a normal distribution. The values of Kurtosis for the models PP-2, Gain-2 and Gain-4 may suggest that the distribution is not normal, but the Kolmogorov-Smirnov Test proves otherwise.

The table below gives the correlations of the ranking measures for each model with one another and two school level variables MOBILITY (MOB) and PMINORITY (PMIN). PMINORITY is included in the models 2 and 4 to demonstrate how efficiently these models remove school level biases, and MOBILITY is included as a control.

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>K-S P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP-1</td>
<td>0.0000017</td>
<td>0.866</td>
<td>0.148</td>
<td>3.170</td>
<td>0.9125</td>
<td></td>
</tr>
<tr>
<td>Gain-1</td>
<td>0.0052139</td>
<td>4.008</td>
<td>0.156</td>
<td>0.207</td>
<td>0.9546</td>
<td></td>
</tr>
<tr>
<td>PP-2</td>
<td>-0.000002</td>
<td>0.703</td>
<td>0.907</td>
<td>0.442</td>
<td>0.9729</td>
<td></td>
</tr>
<tr>
<td>Gain-2</td>
<td>0.0001864</td>
<td>3.135</td>
<td>0.649</td>
<td>0.454</td>
<td>0.5278</td>
<td></td>
</tr>
<tr>
<td>PP-3</td>
<td>-0.0000003</td>
<td>0.788</td>
<td>0.406</td>
<td>0.273</td>
<td>0.8197</td>
<td></td>
</tr>
<tr>
<td>Gain-3</td>
<td>0.0018516</td>
<td>3.679</td>
<td>0.211</td>
<td>0.288</td>
<td>0.7809</td>
<td></td>
</tr>
<tr>
<td>PP-4</td>
<td>-0.0000000</td>
<td>0.096</td>
<td>0.102</td>
<td>0.534</td>
<td>0.9676</td>
<td></td>
</tr>
<tr>
<td>Gain-4</td>
<td>0.0009754</td>
<td>3.180</td>
<td>0.828</td>
<td>0.522</td>
<td>0.6852</td>
<td></td>
</tr>
</tbody>
</table>

From the table above we can observe that the School Rankings obtained from all the models are highly correlated with each other. The correlation range in value from 0.9819 to 0.9883. The correlations in the full model, with both student level and school level variables is 0.9847. Thus we can conclude that each pair of models, for the four different models considered, produces school rankings that are not significantly different from each other.

The last two columns on the table have two school level variables. Of these, PMINORITY was included in the Models PP-2, G-2 and PP-4, G-4. Comparing the correlations, we can conclude that any bias introduced by PMINORITY into the school ranking is removed by including this variable into the model. The Pre-test/Post-test models did a better task of removing the bias than the Gain model since the later models have higher correlations than...
the former models. The school level variable MOBILITY was not included in any of the models and, as can be seen in the table above, it’s bias remains in the school rankings.

Classroom Effective Indices

The process of obtaining Classroom Effective Indices from these models is more complicated than School Effective Indices. All the models above produce regression lines which are school specific, i.e. there is a regression line for each school. If we estimate $r_{ij}$ for the PP models and $r_{ijk}$ for the GAIN models using these lines we obtain student residuals within a specific school. Hence we calculate an average regression line(District Line) and calculate the student residuals from this line with respect to each model as follows.

For the Pre-Test/Post-Test models:

$$r_{ij} = \text{READ96}_{ij} - (\gamma_{00} + \gamma_{1i}\text{READ95}_{ij} + \gamma_{2i}\text{SEX}_{ij} + \cdots)$$ (37)

For the GAIN models:

$$r_{ijk} = \text{READ96}_{ijk} - (\gamma_{00} + \gamma_{1i}\text{READ95}_{ijk})$$ (38)

The following table illustrates the correlations of the student residuals obtained from the eight models.

<table>
<thead>
<tr>
<th></th>
<th>PP-1</th>
<th>G-1</th>
<th>PP-2</th>
<th>G-2</th>
<th>PP-3</th>
<th>G-3</th>
<th>PP-4</th>
<th>G-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP-1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G-1</td>
<td>0.957</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP-2</td>
<td>1.000</td>
<td>0.957</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G-2</td>
<td>0.957</td>
<td>1.000</td>
<td>0.957</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP-3</td>
<td>0.997</td>
<td>0.955</td>
<td>0.996</td>
<td>0.955</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G-3</td>
<td>0.957</td>
<td>1.000</td>
<td>0.957</td>
<td>1.000</td>
<td>0.955</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP-4</td>
<td>0.998</td>
<td>0.567</td>
<td>0.998</td>
<td>0.956</td>
<td>1.000</td>
<td>0.956</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>G-4</td>
<td>0.957</td>
<td>1.000</td>
<td>0.957</td>
<td>1.000</td>
<td>0.955</td>
<td>1.000</td>
<td>0.956</td>
<td>1.000</td>
</tr>
</tbody>
</table>

As can be seen from the above table the Pre-Test/Post-Test Models(PP) and the Gain Models(G) are also highly correlated, with the lowest correlation having a $\rho = 0.955$. We can conclude from the above that the two types of models calculate district-wide residuals with significantly the same distribution.

There is major draw back of calculating district wide residuals from a multi-level model with schools as a level. Within a school, the student residuals, i.e., $r_{ij}$ for $j$ fixed has a normal distribution with a mean of zero. This is not the same for calculating student residuals district wide. Even if the model is not a multi-level model, but a linear regression model, the student residuals will not be normal. This fact should be considered when these residuals are analyzed and utilized.

Though the residuals of the various pairs of models are significantly correlated, this may not be evident from the descriptive statistics below.

7