Decomposition of Estimates in a Layered Value-Added Assessment Model

S. Paul Wright
Dr. William L. Sanders
Value-Added Assessment and Research
SAS Institute Inc.

Abstract. The layered model of Sanders, et al. (1997) is a value-added assessment model that uses multivariate longitudinal mixed model methodology. This model, whose advantages are re-emphasized here, has been criticized for its complexity and lack of transparency. The main purpose of this paper is to provide more transparency for this model by comparing it to three sub-models which provide a decomposition of the layered model estimates. The comparisons make use of simple examples as well as explicit formulas which reveal how the layered model constructs teacher effects from student data. The comparisons reveal how the layered model incorporates past students scores, and future scores when available, to produce estimates with more appealing characteristics than alternative models, particularly in the case when the student population is divided into independent tracks.

1. Introduction

For the analysis of longitudinal student test scores, a number of different models have been proposed and debated ranging from extremely simplistic to complex; see, for example, Lissitz (2005) and Wainer (2004). Advocates of the more complex models have stressed their statistical advantages and the fact that these models have been used successfully for decades in other fields such as agricultural and medical research (Hershberg, 2004; Drury and Doran, 2003; Doran, 2003). Nevertheless, arguments continue to be made that the "transparency" of the method of analysis should be an important criterion in choosing a model (Tekwe, et al., 2004).

The purpose of this paper is to try to increase the transparency of one of the more complex models, the layered teacher model proposed by Sanders, et al. (1997), noting its similarities to and differences from some of the more simple models. Specifically, three alternatives to the layered model, each of which is a sub-model of the layered model, are used to construct an interpretable picture of the layered model estimates. In a sense the sub-model estimates provide a decomposition of the layered model estimates. Many of the ideas presented here appear, to some extent, in McCaffrey, et al. (2004, 2005). The purpose of presenting them here is to provide additional clarity so as to promote a better understanding of value-added models and especially to highlight advantages of the layered model.

Three approaches are used in what follows. First, simple example data sets are used to generate estimates whose properties are, for the most part, easy to see. Second, some general formulas for the estimates are given along with verbal interpretations of what the formulas mean. Third, the matrix equations used to obtain both the estimates and the explanatory formulas are given so that interested readers can conduct their own investigations.

Section 2 briefly presents some general results for the linear mixed model of which the layered model is a particular instance. Results in subsequent sections build on the general formulas of Section 2. Additional technical details for the linear mixed model are given in two appendices. Section 3 discusses the oneway random-effects ANOVA model, the simplest example of a linear mixed model, using this well-understood setting to introduce the concepts of shrinkage estimation and best linear unbiased estimation. Section 4 gives detailed results (i.e., estimates using a particular data set as well as general formulas) for an example with two years of data and two teachers per year. Section 5 presents a more realistic example with three years of data and three teachers per year that fall into two separate tracks. The formulas from the simpler case in Section 4 are seen to carry over to the more complex example in Section 5. Section 6 discusses the layered model, as contrasted with other models, regarding several issues that arise in value-added modeling: the impact of missing values, the question of the persistency of teacher effects, and the issue of omitted variable bias and covariate adjustment.

Throughout this paper, the term "effect" is used to refer to certain terms in a linear model (e.g., fixed effects and random effects in Section 2). This terminology is common even in observational studies and does not necessarily imply causation. For discussion of causal inference see Rubin, Stuart, and Zanutto (2004), Rubin (2005), and Raudenbush (2004).

2. Technical Details, Part I: The Linear Mixed Model

Paradoxically, this attempt to provide more transparency begins with what is the most non-transparent aspect of the model, namely, the model itself. The purpose is to provide the general formulas from which subsequent, more transparent, formulas are derived. The layered teacher model is a special case of a linear mixed model for which the general formula is

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\upsilon} + \boldsymbol{\varepsilon}$ where

$\boldsymbol{\beta}$ is a p-by-1 vector of fixed effects; $\mathbf{X}$ is an n-by-p matrix;
$\boldsymbol{\upsilon}$ is a q-by-1 vector of random effects; $\mathbf{Z}$ is an n-by-q matrix;
$E(\boldsymbol{\upsilon}) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{\upsilon}) = \mathbf{G}$; $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{\varepsilon}) = \mathbf{R}$; $\mathrm{Cov}(\boldsymbol{\upsilon}, \boldsymbol{\varepsilon}) = \mathbf{0}$.
$\mathbf{V} = \mathrm{Var}(\mathbf{y}) = \mathrm{Var}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathrm{Var}(\mathbf{Z}\boldsymbol{\upsilon} + \boldsymbol{\varepsilon}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^{T} + \mathbf{R}$.

The fixed effects ($\boldsymbol{\beta}$) are estimated by generalized least squares using

$$\mathbf{b} = (\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{X})^{-1}\,\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{y} = \mathbf{Q}_1\,\mathbf{y}.$$

The random effects ($\boldsymbol{\upsilon}$) are estimated using

$$\mathbf{u} = (\mathbf{Z}^{T}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\,\mathbf{Z}^{T}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{G}\mathbf{Z}^{T}\mathbf{V}^{-1}\mathbf{r} = \mathbf{Q}_2\,\mathbf{r}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}\mathbf{b}$. Subsequently, the values in $\mathbf{r}$ will be referred to as "deviation scores" or simply "deviations." Later it will be seen that the elements of $\mathbf{Q}_1$ are useful in understanding $\mathbf{b}$ and the elements of $\mathbf{Q}_2$ are useful in interpreting $\mathbf{u}$.

The examples in this paper use test scores from a single cohort of students in a single subject over several years under different sequences of teachers. The $\mathbf{y}$-vector contains the scores. The columns of $\mathbf{X}$ indicate the year of each score, one column for each year, so that the $\boldsymbol{\beta}$-vector contains an overall mean score for each year. The columns of Z identify the teacher(s) associated with each score so that the $\boldsymbol{\upsilon}$-vector contains teacher effects. When the pieces are put together, the elements of $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\upsilon}$ are teacher-level mean scores, and the elements of the $\boldsymbol{\varepsilon}$-vector are residuals (sometimes called random "errors"), i.e., deviations from the teacher-level means.

The $\mathbf{G}$ matrix contains teacher variance components; these measure how much difference in effectiveness there is among the teachers each year. It is customary to assume that each teacher's effectiveness is independent of any other teacher so that $\mathbf{G}$ is a diagonal matrix. In this paper, for simplicity and clarity of exposition, all the (diagonal) elements of **G** are assumed to have the same value, τ^2 (tau-squared).

The **R** matrix contains the variances and covariances among the residuals (elements of $\boldsymbol{\varepsilon}$). Since students are assumed to be independent, but residuals belonging to the same student are *not* independent, the **R** matrix is block-diagonal with a block for each student. It is customary to assume that every student's block of **R** is the same (this is analogous to the assumption of equal variances in regression analysis). For simplicity and clarity, it will also be assumed that the diagonal elements of **R** all have the same value σ^2 (sigma-squared). In such a case, σ^2 can be factored out so that $\mathbf{R} = \sigma^2 \mathbf{R}_0$ where $\mathbf{R}_0$ is a block-diagonal *correlation* matrix. To further simplify the examples, all the correlations in $\mathbf{R}_0$ are given the same value ρ (rho), and the value of rho will be either $\rho=0$ or $\rho=0.7$.

Finally, let $\theta = (\tau^2 / \sigma^2)$ indicate the ratio of the teacher variance to the residual variance. Then,

$$\mathbf{V} = \mathbf{ZGZ}^T + \mathbf{R} = \sigma^2 (\theta\, \mathbf{ZZ}^T + \mathbf{R}_0) = \sigma^2 \mathbf{V}_0.$$

The formulas given earlier for the estimated fixed and random effects (**b** and **u**) produce the same results if $\mathbf{V}_0$ is substituted for **V**, $\mathbf{R}_0$ is substituted for **R**, and **G** is replaced by a diagonal matrix with (diagonal) elements θ. To simplify notation, subsequent references to **V**, **R** and **G** are assumed to refer to $\mathbf{V}_0$, $\mathbf{R}_0$ and the diagonal matrix with elements θ.

In practice, the variance and covariance components in **R** and **G** (almost always) must be estimated from the data simultaneously with **b** and **u**. Since the purpose of this paper is to clarify the interpretation of the **b** and, especially, **u**, given **R** and **G**, variance component estimation is not discussed; see Searle, et al. (1992) for more information.

3. Example #1: A Oneway Random-Effects ANOVA

In general, the examples in this paper involve more than one score for each student. The current example is the lone exception, using only a single year of data with test scores for a group of students under three different teachers identified as A, B, and C. This corresponds to the simplest possible mixed model analysis, a oneway random-effects analysis of variance. This example is used to demonstrate two important concepts that are important for understanding mixed model estimates, the concepts of shrinkage estimation and of best linear unbiased estimation of fixed effects.

Table 1 displays a very simple example with one student for each teacher. The "**X**" column is the **X** matrix of the linear mixed model described in Section 2. In this example it is simply a column of ones, and the corresponding vector of estimated fixed effects (**b**) consists of a single element, the estimated overall mean score. The three "**Z**" columns make up the **Z** matrix of the mixed model. Each column identifies the teacher of the student whose score is in that row. The "$\mathbf{Q}_1$" column contains the elements of the $\mathbf{Q}_1$ matrix (in this case a vector) which shows the weight given to each student's score in the estimation of the overall mean. In this example, each score gets equal weight so that the estimated overall mean is just the arithmetic average of the scores, 400. The "**Dev**" column contains the deviation scores (**r**), that is, the student scores minus the estimated overall mean of 400. The "$\mathbf{Q}_2$" columns (one column for each teacher) contain the elements of the transposed $\mathbf{Q}_2$ matrix which contains the weights by which the deviation scores are multiplied to obtain the estimated teacher effects (**u**). In this case, each teacher effect is equal to the deviation score of that teacher's student. The "Est. Mean" column shows the estimated mean score for each teacher which is the estimated overall mean (400) plus the teacher effect (from column "**u**").

The weights in $\mathbf{Q}_1$ and $\mathbf{Q}_2$ depend on the values in the **R** and **G** matrices. In this example with only one year of data, the **R** matrix is a diagonal matrix with each (diagonal) element equal to one. The results in Table 1 are for a **G** matrix with diagonal elements $\theta = 1000$, i.e., the variation among teachers is extremely large compared to the variation among students having the same teacher.

For comparison, Table 2 contains results when θ is equal to one (1). The teacher effects are now only half as large as before; this is due to "shrinkage estimation" or (better) "best linear unbiased prediction" (BLUP). The estimated means are also shrunk toward the estimated overall mean of 400. See Raudenbush and Bryk (2002, pp. 45-51, 157-158) or Robinson (1991) for additional details and justification concerning BLUP, including an empirical Bayesian justification. In the case of the oneway ANOVA model, the amount of shrinkage can be expressed by the formula

$u_j = w_j \cdot \text{mean}(\mathbf{r}_j)$ where

u_j is the j-th element of **u**, the estimated teacher effect for the j-th teacher,
mean($\mathbf{r}_j$) is the arithmetic mean of the deviation scores for that teacher's students,
w_j is the shrinkage factor: $w_j = (n_j\,\theta) / (n_j\,\theta + 1)$ where
θ is the ratio of teacher variance to error variance, the diagonal element of **G**, and
n_j is the number of students taught by the j-th teacher.

In Tables 1 and 2 (also in Tables 3 and 4), the non-zero values in the $\mathbf{Q}_2$ columns are equal to the shrinkage factor w_j. In Table 1, $n_j = 1$ and $\theta = 1000$ so that w_j is essentially equal to one (no shrinkage). In Table 2, $n_j = 1$ and $\theta = 1$ so that $w_j = 0.5$. Note that since the value of w_j is a function of the product ($n_j\,\theta$), the amount of shrinkage is equally affected either by a change in θ or a change in n_j. For the examples, it was more convenient to manipulate θ in order to have small samples for display in the tables. In actual practice, θ is a population parameter with some fixed but unknown value that must be estimated, and the estimates in **u** for different teachers receive different amounts of shrinkage primarily due to differences in n_j. Appendix #1 discusses how varying values of n_j can be incorporated into the calculations in a way that is both computationally efficient and increases transparency.

Tables 3 and 4 show the consequences of different teachers having different numbers of students, without and with shrinkage ($\theta = 1000$ and $\theta = 1$, respectively). Table 3 (no shrinkage) looks just like Table 1 while Table 4 shows how the amount of shrinkage decreases as n_j increases. The "All" row shows that the observed mean of all scores is now 412.5, rather than 400, while the estimated overall mean is 400 in Table 3 and 404.17 in Table 4. Recall that the estimated overall mean score is obtained by multiplying the values in the "Score" (or "Mean Score") column by the $\mathbf{Q}_1$ values and summing those products. In general, the estimated overall mean in the oneway random-effects ANOVA model (in all cases including Tables 1 and 2) is a *weighted* average of the observed teacher means ("Mean Score" column), and the weights are the same w_j values that are used in shrinkage estimation of the teacher effects. The values in $\mathbf{Q}_1$ are just the w_j values rescaled so that they sum to one. This choice of weights produces an estimated overall mean which is a "best linear unbiased estimate" (BLUE for short); see Searle, et al. (1992, section 3.3 and appendix S.2) or Raudenbush and Bryk (2002, pp. 38-45, the "unique, minimum variance, unbiased estimator") for more information. Depending on the amount of shrinkage, the estimated overall mean varies from an unweighted mean of the observed teacher means (Table 3, no shrinkage) to the simple mean of all student scores (as θ approaches zero).

4. Example #2

The remaining examples in this paper use test scores from a single cohort of students in a single subject over several years under different sequences of teachers. The scores are assumed to be scaled such that score differences from year to year ("gains") are meaningful. Table 5 contains a simple example consisting of four students over two years with two teachers each year.

The two columns of **X** (**X1** and **X2**) contain zero/one indicator variables identifying the year. Consequently, the fixed effects estimates (the two rows of **b** in Table 6) contain the estimated overall mean scores for each year, and the deviation scores (**r**, see Tables 7 and 8) contain deviations from those yearly means. Two sets of **Z**-matrix columns are shown; each set has four columns, one for each teacher. The columns labeled "**Z (NLM)**" represent a non-layered model: for each observation (each row of **Z**) there is a single "1" in the column identifying the teacher who taught that student that year. In effect, the non-layered model says that a student's score each year is made up of an overall mean (from **β**) plus the effect of that year's teacher (from **υ**) plus some random residual variation (from **ε**). This model seems unrealistic in that it assumes that each teacher starts with a set of average students each year, ignoring the effects of instruction in earlier years. Even if all four students in this example started out at the same achievement level (e.g., all were initially average), by the end of the first year, teacher B's students are clearly ahead of teacher A's students; and it is reasonable to assume that students of teacher B will still be ahead of teacher A's students when they begin year two. Thus, the effect of the second year teacher should represent what they add on to the result of the first year's instruction. This is the basic rationale for value-added assessment.

Strictly speaking, the effect for a year-one teacher (e.g., teacher A) cannot be a value-added effect since there are no scores from previous years to use as the basis for determining how much value has been added. Consequently, results for teacher A are not directly relevant in a discussion of value-added modeling. They are included in this example for pedagogical reasons because elements of the formulas for teacher A's effect will be found embedded in the formulas for value-added effects in Example #3.

The columns in Table 5 labeled "**Z (LM)**" represent the layered model. Each row of **Z** for a second year score now contains two "1"s, one to identify the first year teacher and one to identify the second year teacher. This model is saying that a student's score each year is made up of an overall mean plus the *cumulative* effects of all the teachers that the student has had so far, plus some random residual variation. In the examples that follow, the layered model will be seen to have a number of advantages which are not immediately apparent from this simple description.

In order to calculate the estimates **b** and **u** using the formulas given in Section 2, a **V** matrix is needed, thus **R** and **G** matrices are needed (in practice they are usually estimated from the data). As in Example #1, **G** is a diagonal matrix in which, for simplicity, all the (diagonal) elements have the same value θ. For easier interpretability, θ = 1000 was chosen to minimize the amount of shrinkage in the random effect estimates (**u**).

Throughout the remainder of the paper estimates (mainly **u**, sometimes **b**) from four different models are compared: [1] the non-layered model (NLM) with correlations of zero (ρ=0) in **R**, abbreviated NLM(0), [2] the non-layered model with ρ=0.7 (a value commonly found in practice), abbreviated NLM(0.7), [3] the layered model (LM) with ρ=0, abbreviated LM(0), [4] the layered model with ρ=0.7, abbreviated LM(0.7). Model [4] represents the layered teacher model which is the main focus of this paper. The other models have been included primarily for their usefulness in explaining the layered model, but some of these models have also been used in practice (see the Discussion). In certain circumstances, the results from these models correspond to what would be obtained from a simpler analysis, but in general this is not the case.

The estimates from the Example #2 data set (Table 5) are shown in Table 6. This is not a particularly informative illustration since most of the estimates are the same in all four models. It has been included because of its simplicity and so that the contents of the **X** and **Z** matrices could be presented easily. A more revealing example is given later. First, however, formulas will be developed that show how student (deviation) scores get combined to produce these estimates.

The fixed effects estimates (**b**) are straightforward: each element of **b** is the arithmetic mean of the four student scores for the appropriate year. The formula is the same for all four models and is therefore omitted. The random effect estimates (**u**), despite the equality of their values for different models in this example, come from quite different formulas. Tables 7 and 8 show the elements of $\mathbf{Q}_2$ (transposed) for each model for a year-one teacher (teacher A) and a year-two teacher (teacher C), respectively. The NLM(0) column, for example, contains the elements of the *row* of $\mathbf{Q}_2$ corresponding to teacher A (or C). The NLM(0) teacher effect for teacher A (or C) is obtained by multiplying the elements of the "Deviation" column by the corresponding elements of the "NLM(0)" column and adding up the resulting products.

Concise, transparent formulas for the estimates from the various models are given in Table 9 using the following notation. Let "M" be the mean of the *deviation scores* for a set of students. An integer suffixed to "M" identifies the year: M1 is the mean of year-one deviations, M2 is for year two. Subscripts identify the teacher(s) whose students make up the mean.

NLM(0). The interpretation of the values for the NLM(0) estimates is straightforward: each teacher effect is simply the average of the deviation scores of the two students that each teacher taught. Using the notation given above, the formula for the NLM(0) teacher effect for teacher A is simply $M1_A$, the mean of the year-one deviation scores for the two students of teacher A, with a value of −20 (the mean of −25 and −15). For teacher C the formula is $M2_C$, the mean of the year-two deviation scores for students of teacher C, also with a value of −20 (the mean of −45 and +5).

NLM(0.7). The NLM(0.7) estimates, using the Table 9 formulas, are

Teacher A: $M1_A - 0.70\,(M2_A - M2_{AB}) = (-20) - 0.70\,[(-20) - 0] = -6;$

Teacher C: $M2_C - 0.70\,(M1_C - M1_{CD}) = (-20) - 0.70\,[0 - 0] = -20.$

In words (using teacher C as an example), $M2_C$ is the mean of the year-two deviation scores for students of teacher C (−45 and +5), $M1_C$ is the mean of the year-one deviations for those same students (−25, +25) while $M1_D$ is the mean of the year-one deviations for students who had teacher D, rather than teacher C, in year two (−15, +15), so that $M1_{CD}$ is the mean of the year-one deviations for students who had either teacher C or teacher D in year two (which, in this example, is all four students). The effect for teacher C is an adjusted mean of the deviation scores of teacher C's students, adjusted based on their previous-year deviation scores. The formula is precisely that of an adjusted mean in an analysis of covariance having Teacher as the grouping variable and the year-one deviation score as the covariate. The value of 0.70 in the formula is the pooled within-teacher regression coefficient (which in this case is the same as the correlation coefficient in the **R** matrix) for predicting the year-two deviation score from the year-one deviation score.

The effect for teacher A has a similar interpretation. In the case of teacher A, however, the response variable is the year-one deviation score and the covariate is the year-two deviation score. As a general rule, in traditional analysis of covariance, a future score is not used to adjust an earlier score. In the analysis of covariance literature (e.g., Urquhart, 1982), this is the problem of having a covariate (the future score) which is potentially affected by the treatment (the current year teacher). In such a case the standard analysis, as represented by the NLM(0.7) formula, is generally inappropriate. The layered model employs a very different approach to incorporating future scores; see the LM(0) section below.

In general, a teacher effect can be thought of as an average of the observed deviation scores minus the expected deviation scores of that teacher's students. The technical details behind this assertion are given in Appendix #2, but this interpretation is clearly evident in the NLM(0.7) estimates. For teacher C, for example, $M2_C$ is the mean observed deviation score and $0.70\times(M1_C - M1_{CD})$ represents the mean expected deviation score from a regression on the students' previous-year deviation scores.

LM(0). The LM(0) estimates are

$$\begin{aligned}\text{Teacher A: } & M1_A + 0.5\,[(M2_A - M2_{AB}) - (M1_A - M1_{AB})] \\ & = (-20) + 0.5[(-20 - 0) - (-20 - 0)] = -20;\end{aligned}$$

$$\text{Teacher C: } M2_C - M1_{AB} = (-20) - 0 = -20.$$

In the case of NLM(0.7) above, the formulas for Teacher A and Teacher C had the same interpretation. This is not the case with LM(0): here the formula for Teacher C looks quite different from that for Teacher A. The Teacher C formula is the simpler to explain of the two. Like the NLM(0.7) estimate, the LM(0) estimate for teacher C is an average observed minus expected deviation score. In this case the average expected deviation score is just the mean of the previous-year deviation scores of *all students* ($M1_{AB}$ which happens to equal zero). This makes an important point about certain differences among the models. In general, the observed responses (the values in the **y** vector of the linear mixed model) are correlated. These correlations are expressed in the **V** matrix. It is these correlations that make an expected deviation score possible. The **R** matrix (with non-zero off-diagonals) contains within-student correlations; this allows an expected deviation score for a student to be based on other deviation scores *by the same student*. This was the case with the NLM(0.7) estimates.

Random effects in the model (expressed in the **Z** and **G** matrices via the $\mathbf{ZGZ}^T$ component of **V**) induce between-student correlations; in this case an expected deviation score for a student is based on deviation scores *by other students who have shared the same teacher(s)*. In the layered model, because of the layering, students who shared the same year-one teachers with teacher C's students enter the expected deviation score calculation, producing the $M1_{AB}$ term since teacher C received equal numbers of students from both teachers A and B. This is in contrast to the NLM(0) estimate for teacher C: without layering there is no correlation between year-two and year-one deviation scores and thus no basis for obtaining an expected deviation score. Consequently, the NLM(0) estimate is determined entirely by the observed deviation scores, with no adjustment.

The LM(0) formula for teacher A looks more formidable; indeed, it does not look like an observed minus expected formula at all. While it is possible to write the formula in an observed minus expected format, the alternative formula in Table 9 seems more transparent. This formula, like the NLM(0.7) formula for teacher A, uses *future* scores to adjust the teacher effect, but the contrast could not be more striking. In the NLM(0.7) case, “better than average” year-two scores ($M2_A - M2_{AB} > 0$ in the formula) cause teacher A’s effect to be lower. In contrast, in the LM(0) formula, relatively better future performance [$(M2_A - M2_{AB}) - (M1_A - M1_{AB}) > 0$] causes teacher A’s effect to be higher! More will be said about this in the Discussion section.

LM(0.7). Finally, the LM(0.7) estimates are

Teacher A: same as LM(0);

Teacher C: $= M2_C - M1_{AB} - 0.7\,(M1_C - M1_{CD}) = (-20) - 0 - 0.7\,(0 - 0) = -20.$

The effect for teacher C is, again, an average observed minus expected deviation score. The expected part of the formula combines the expected terms from the NLM(0.7) and the LM(0) formulas. That is, only the LM(0.7) model makes use of both the within-student and the between-student correlations.

Careful attention to what happens to the deviation scores when entered into the above formulas reveals why this example is rather uninformative. Because of the balanced nature of the assignment of students to teachers (with teacher A’s and B’s students equally distributed to teachers C and D), the deviation scores in the expected part of the formulas tend to cancel one another out except in the NLM(0.7) case for teachers A and B.

5. Example #3

In this example, attention is restricted to teachers other than year-one teachers. This is because the focus of attention in this paper is the layered teacher model and related value-added models in which cases year-one teacher effects are not usually reported. As previously noted, year-one teacher effects are not value-added effects.

Table 10 contains an example with three years of scores and three teachers per year. The year-one teachers are A, B, C; the year-two teachers are D, E, F; the year-three teachers are G, H, J. Note, however, that the three teachers per year fall into two tracks (called strata by McCaffrey, et al., 2004, 2005). In the first track, students from teachers A and B progress to teachers D and E then to teachers G and H. In the second track, students have teacher C, then teacher F, then teacher J.

Table 10 also contains the elements of $\mathbf{Q}_1$, the weights given to individual student scores (the “Score” column) to obtain the three yearly fixed-effect means in vector **b** (see Table 11). Notice that the weights are not equal, that is, each estimated yearly overall mean is not simply the average of all the student scores. Instead it is the mean of the teacher means as described in Section 3 (the Table 3 case, no shrinkage). The “Mean Score” column of Table 11 shows the mean student score for each teacher as well as the simple mean of all scores for each year.

Table 11 also shows the estimated random effects from each model. Also shown, in the last column, are the means of the student gain scores (current score minus previous score) for each year and for each teacher. The formulas for the estimated teacher effects are discussed in more detail below, but certain features are apparent in Table 11. First, as in the previous example, each NLM(0) teacher effect is simply the mean deviation score (the “Mean Score” for the teacher minus the fixed-effect mean for that year). Second, the NLM(0.7) estimates resemble the NLM(0) estimates and do not appear to be indicating value-added as reflected in the “Mean

Gain" column. More insight into all these estimates can be gained by examining the elements of $\mathbf{Q}_2$.

Tables 12 and 13 contain the weights from $\mathbf{Q}_2$ for producing the estimated teacher effects for two second year teachers, D and F (one from each track), and two third year teachers (G and J). The "Dev" column shows the deviation scores. These are multiplied by the elements of $\mathbf{Q}_2$ to get the estimated teacher effects. These deviation scores are obtained by subtracting from each student's score the fixed-effect mean for the appropriate year (300, 400, or 500).

The first thing to notice about the weights in $\mathbf{Q}_2$ is that the two tracks are independent. Effects for teachers in the first track (A, B; D, E; G, H) depend only on deviation scores of students in that track; the same is true for teachers in the second track (C; F; J).

Formulas derived from these weights are given in Table 14 for year-two teacher D and year-three teacher G. In the NLM(0.7) and LM(0.7) formulas the symbol "b" appears. This is the pooled within-teacher multiple regression coefficient for predicting the year-three deviation score from the year-one and year-two deviation scores (or for predicting year two from years one and three, or year one from years two and three). Given a compound symmetric correlation matrix (**R**) with a constant off-diagonal correlations of "r" (r = 0.7 in this example), it can be shown that the value of "b," in the case of a 3-by-3 matrix for each student, is $b = r / (1 + r) = 0.411765$. (More generally, for an m-by-m compound symmetric correlation matrix, $b = r / [1 + (m-2) r]$.) The value "0.0515," which appears repeatedly in Tables 12 and 13, is just (b/8).

No formulas are given for the effects of the teachers in the second track (teachers F and J) for two reasons. First, the interpretations of these effects are readily apparent from the elements of $\mathbf{Q}_2$ in Tables 12 and 13. Specifically, in the non-layered models, both NLM(0) and NLM(0.7), the teacher effects are simply mean deviation scores. In both of the layered models, the teacher effects are mean gain scores. Second, the formulas are actually the same as for teachers D and G (with appropriate changes in subscripts), but certain simplifications occur because there is only one teacher per year in this track. These simplifications are described in the appropriate paragraphs below.

NLM(0). The interpretation of the NLM(0) teacher effects is exactly the same as in the previous example: it is the (unadjusted) mean deviation score of the teacher's students.

NLM(0.7). For teachers D and G in track one the formulas from Table 14 are

Teacher D: $M2_D - b\,(M1_D - M1_{DE}) - b\,(M3_D - M3_{DE})$.

Teacher G: $M3_G - b\,(M2_G - M2_{GH}) - b\,(M1_G - M1_{GH})$.

The interpretation of the NLM(0.7) teacher effects is the same as in the previous example: it is the adjusted mean (as in analysis of covariance) of the deviation scores. For the year-two teacher D, this involves the use of a future score as a covariate as discussed in Example #2. However, there is an important difference here from the previous example. The means of the covariates, to which the teacher mean deviations are adjusted, are not grand means (over all students) as are usually used in analysis of covariance; instead they are the mean deviation scores of students *in the same track* (e.g., $M1_{GH}$ and $M2_{GH}$ for teacher G). This has the undesirable consequence of causing the second track teachers (F and J), whose student gains were only average but whose scale scores were high, to have very large positive teacher effects, while first track teachers E and H, whose student gains were the highest, have much lower teacher effects. Thus, the NLM(0.7) effects for teachers D and E are subject to two potentially undesirable influences: the adjustment of past scores using future scores and the within-track centering of the covariates.

For teachers F and J in the second track, it turns out that there is no adjustment. Because there is only one teacher each year in this track, the within-track mean of each covariate is the same as the teacher mean; i.e., the difference between the teacher mean and the track mean is always zero. For example, the ($M1_D - M1_{DE}$) term in the teacher D formula becomes ($M1_F - M1_F$) = 0 for teacher F. Consequently, the NLM(0.7) effects for teachers F and J are not value-added effects at all since they do not take into account the previous achievement levels of these teachers' students.

LM(0). For teachers D and G in track one the formulas from Table 14 are

Teacher D: $M2_D - M1_{DE} + .5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})]$.

Teacher G: $M3_G - M2_{GH}$.

For year-three teacher G, the formula is the same as for the year-two teachers in Example #2. In both instances, the teacher is a terminal-year teacher; no future data is available for that teacher's students. In both instances the mean deviation score of the teacher's students is adjusted by subtracting off the mean of the previous-year deviation scores of "all students." In this example, however, it is seen that "all students" actually refers only to students *in the same track*. In the NLM(0.7) case such within-track adjustments disguised the between-track differences, resulting in teacher effects that did not reflect the value added by the teacher. In the LM(0) case, this problem does not occur for reasons that are most easily explained in connection with the LM(0.7) model which is discussed below.

Students of year-two teacher D, unlike teacher G, have both previous-year and next-year deviation scores. The previous year scores are handled exactly as they were for teacher G: the mean previous year deviation score of students in the same track is subtracted off. The next-year deviation scores are handled as they were for the year-one teachers in the previous example (see the discussion of the LM(0) effect for teacher A in Example #2).

For teachers F and J in the second track certain simplifications occur (and no formulas are shown in Table 14) because there is only one teacher each year in that track. For year-three teacher J, for example, the previous year mean deviation for all-students-in-the-same-track is just the previous-year mean deviation of teacher J's students. Thus, the ($M3_G - M2_{GH}$) for teacher G becomes ($M3_J - M2_J$) for teacher J. That is, the teacher effect for teacher J is a mean gain. Similarly, for year-two teacher F the adjustment for past scores becomes ($M2_F - M1_F$). In addition, the adjustment for future scores is zero because the $[(M3_D - M3_{DE}) - (M2_D - M2_{DE})]$ term in teacher D's formula in Table 14 becomes $[(M3_F - M3_F) - (M2_F - M2_F)] = 0$.

LM(0.7). For teachers D and G in track one the formulas from Table 14 are

Teacher D: $M2_D - M1_{DE} - 0.7\ (M1_D - M1_{DE}) + .5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})]$.

Teacher G: $M3_G - M2_{GH} - b\ (M2_G - M2_{GH}) - b\ (M1_G - M1_{GH})$.

As in Example #2, the LM(0.7) estimated teacher effects combine the adjustments of the NLM(0.7) and the LM(0) models. For example, the LM(0.7) formula for teacher G is the NLM(0.7) formula with the addition of the $M2_{GH}$ term which is the between-students adjustment for past scores from LM(0). Because $M2_{GH}$ is a mean of deviation scores, it contains the estimated mean *scale score* for track one students in year two compared to the estimated mean *scale score* of *all* students in year-two. In other words, the $M2_{GH}$ term *reintroduces the between-track differences* that are missing from the NLM(0.7) estimates. For teacher D, the added $M1_{DE}$ term serves this function. Also, for teacher D, the NLM(0.7)-type adjustment for future scores is replaced by the LM(0)-type adjustment which is described in Section 4 and discussed further in Section 6.

For teacher F and J in the second track, the simplifications described under NLM(0.7) and LM(0) both apply with the result that the teacher effects represent mean gains.

6. Conclusion and Discussion

The Four Models as a Decomposition. The goal of this paper has been to present examples, formulas, and verbal interpretations to demystify the layered teacher model, and to contrast that model with selected other models. The other models were chosen primarily for their usefulness in decomposing the estimates from the layered model. Thus, the NLM(0) model represents an unadjusted analysis with teacher effects that simply reflect the students' scores. In this respect it resembles the type of assessment mandated by the NCLB legislation. The NLM(0.7) and LM(0) models adjust the teacher effects, in two very different ways, to account for both past and future scores. The NLM(0.7) model uses a within-student adjustment comparable to an analysis of covariance. The LM(0) model uses a between-students adjustment. The layered teacher model, represented by LM(0.7), employs both of these adjustments.

Discussion: Missing Data. One characteristic that is shared by all four of the models used in this paper is the use of the linear mixed model formulation. An important advantage of this formulation is that it accommodates missing data so that all student scores can be used, including those of students who were not present every year. This contrasts with other popular and simpler approaches to value-added modeling, such as the analysis of gain scores (with or without additional covariates) and the covariate adjustment model in which present scores are regressed on past scores (with or without additional covariates). In both of these approaches, missing student scores can result in a troubling loss of data, especially in districts with highly mobile student populations. The resulting estimates from these simpler models can be seriously biased unless the missing data process is MCAR (missing completely at random: the probability that a score is missing does not depend on any data values, either observed or missing; see Little and Rubin, 2002). In contrast, the linear mixed model requires the less stringent assumption of MAR (missing at random: the probability depends only on observed data values, not on unobserved values). In reality, it is likely that missing values are, at least to some extent, NMAR (not missing at random: the probability depends on unobserved data values) in which case the linear mixed model estimates may be biased. However, this bias can often be minimized by including more observed scores in the model (Collins, et al., 2001). In actual practice the layered model typically includes multiple years of scores in multiple subjects. Consequently the risk of bias due to NMAR should be considerably less than with the simpler models.

Because missing data are ubiquitous in the analysis of longitudinal student test scores, the linear mixed model method of handling missing data is a major advantage. However, it is worth noting that, as shown by McCaffrey, et al. (2005), even when there are no missing values, the linear mixed model formulation ("joint modeling" of scores in the terminology of McCaffrey, et al., 2005) usually produces better estimates (smaller mean squared error) than the simpler approaches.

Discussion: Persistency of Teacher Effects. As mentioned above, the four models used in this paper were chosen primarily for their usefulness in explaining the layered model, but to what extent are these four models actually used in practice? The LM(0.7) model is the layered teacher model of Sanders, et al. (1997) which is being used in Tennessee and several districts in other states (via services provided by SAS Institute Inc.). The NLM(0) model, as mentioned, resembles the type of assessment mandated by NCLB which is also, obviously, widely used. To this author's knowledge, the LM(0) model has never actually been used in practice; it was included here purely for expository purposes.

Similarly, the NLM(0.7) model has not been used in practice. However, one of the recent areas of debate concerning value-added assessment is essentially NLM(0.7) versus LM(0.7). McCaffrey, et al. (2004, 2005) present a general model for longitudinal student outcomes which includes, as special cases, the models discussed above along with a number of additional models. A unique feature of their general model is the incorporation of persistency parameters which indicate the degree to which, *on average*, teacher effects carry over from one year to the next. In effect, this model allows those elements of the **Z** matrix which represent the effect of a teacher on *future* student performance to vary between zero and one. Setting all the persistency parameters to zero results in a non-layered model; setting them all to one produces a layered model. In terms of the models discussed above (assuming non-zero correlations in the **R** matrix), the persistency model should produce results that are a compromise between the NLM(0.7) and the LM(0.7) models. Realistically, one might anticipate persistency values slightly less than one but considerably larger than zero.

In McCaffrey, et al. (2004) the persistency model was applied to 678 students in five selected low-poverty elementary schools from a suburban school district. McCaffrey, et al. (2005) report results from applying the model to 9295 students in a large urban school district. Remarkably, in both instances, the estimated persistency parameters were quite small, 0.3 at the most and often much less. Mathematically, persistency parameters near zero suggest that the best fitting model is very close to the NLM(0.7) model. Unfortunately, as shown in Example #3, the NLM(0.7) estimates can exhibit undesirable behavior. This is especially evident in Table 11 where the NLM(0.7) estimates behave similarly to the unadjusted NLM(0) estimates and do not seem to be indicating value-added at all.

McCaffrey, et al. (2004, 2005) indicate that the persistency model is more susceptible to bias from omitted variables than is the layered model. Specifically, they indicate that while *both* layered and non-layered models (and the general model with low persistency as well) are susceptible to bias from omitted variables that are correlated with student *growth in achievement*, only the non-layered (and low persistency) model is susceptible to bias from omitted variables that are correlated with the *level of student achievement*. It is essentially universally true that the magnitudes of the correlations of omitted variables (typically demographic and contextual variables) with achievement *level* are much larger than with achievement *growth*. Thus the risk of bias is much greater in non-layered (and low persistency) models.

On the other hand, what the examples in this paper have shown is that even in the absence of omitted variables, the NLM(0.7) estimates exhibit troubling behavior. First, the manner in which this model uses future scores to adjust teacher effects seems inappropriate since it penalizes teachers whose students go on to perform "unexpectedly well" in future years. This type of adjustment is appropriate only if teacher effects do not persist into the future (which, of course, is what a non-layered model assumes). Perhaps more troubling is that when students fall into tracks or strata (as is often the case in practice), the NLM(0.7) estimates fail to account for between-track differences, resulting in estimates that may fail to reflect value added. Consequently, even though the low persistency general model may "fit" better than the layered model, the risks involved in getting that "better fit" make this model less attractive for value-added analysis.

Discussion: Covariate Adjustment. A common feature of many proposed value-added models, especially those that use the hierarchical linear modeling approach, is the inclusion of covariates to adjust explicitly for "non-instructional" variables that are thought to affect academic achievement (Webster, 2005; Kiplinger, 2004). This is one way that researchers have attempted to address the problem of omitted variables bias mentioned above. Such covariates typically include student characteristics (race/ethnicity, poverty status, etc.) measured at the individual student level and/or at some higher level of grouping (classroom or school ethnic composition, average poverty status, etc.). As the above examples have shown, the LM(0.7) model also contains adjustments both at the student-level (via the **R** matrix) and at the teacher/classroom level (via layering). These layered model adjustments make use of student scores only, without the explicit inclusion of non-instructional variables. One danger in using explicit covariate adjustment with non-instructional variables is the well-documented risk of overcorrection (McCaffrey, et al., 2005, pp. 116-117; McCaffrey, et al., 2004, pp. 86, 95; Ballou, et al., 2004, pp. 38-39) which could hide other factors which could negatively affect student academic performance. For example, if schools serving high concentrations of poor and/or minority students tend to get assigned a disproportionate number of beginning or underqualified teachers, and these teachers tend to be less effective; then adjusting for poverty and/or ethnicity at the classroom level could lead to erroneous estimates of the teacher effects, in this case making the teachers appear to be more effective than they actually are.

Another aspect of the covariate adjustment issue is the distinction between covariates whose values endure over time (time-invariant covariates) and those which change with time (time-varying covariates). Many of the covariates that are included in covariate adjustment models are enduring characteristics of the student, classroom, or school. For example, a student who was in a high-percent-poverty school this year was probably in a similar (or the same) school last year, and the effect of that milieu on the student's achievement level was probably similar both years, so that the effect on the student's *growth* in achievement would be negligible. As noted above, the omission of such covariates does not bias the estimates from the layered model, but it does bias estimates from non-layered models.

Of more concern are factors that affect only one year's score and which, consequently, can bias both layered and non-layered models. Imagine that some known or unknown external pulse unrelated to the teacher has caused a year-three teacher's students to score higher (or lower) in his/her classroom during the year of instruction. One example of such a pulse would be a local disaster or crisis at annual testing time that caused students to perform uncharacteristically poorly. Another example would be if students received unauthorized assistance on the test, possibly from the teacher (i.e., cheating occurred). Such a pulse will create an upward or downward bias in the teacher's estimated effect when year three is the final year of data. Now suppose that in the following year (year four) these same students had academic performance considerably different than the previous biased effect would indicate and more consistent with what would be expected without the external pulse. As the examples have shown, the layered model that includes year four data will make an adjustment for those future scores, producing a "revised" year-three estimated teacher effect for that teacher that corrects for the upward or downward bias produced by the external pulse in year three. Comparing the initial estimates from the year-three model to the revised estimates from the year-four model provides the opportunity to identify instances where an external pulse may have occurred.

Jointly modeling all four years but without the layering, as in NLM(0.7), or with low persistency, will also produce a revised year-three estimate; but in this case the bias will be exaggerated rather than reduced. As shown in Example #2, a year-three teacher whose students perform unexpectedly well (or unexpectedly poorly) in year four as compared to year three will have their teacher effect adjusted downward (upward). Of course, simpler models that do not jointly model multiple years of data (traditional gain models and covariate adjustment models) provide no opportunity to revise estimates based on future data.

To summarize: In most circumstances, exogenous non-instructional covariates are probably not needed or desirable in the layered value-added model. However, if it is required that they be included as a matter of policy (Webster, 2005), then the general form of the linear mixed model can accommodate their inclusion.

Summary: Advantages of the Layered Model. The layered model has been the focus of this paper because (1) it has been widely used in practice, (2) it is a complex model which has been criticized for its complexity and lack of transparency, (3) it has a number of advantages relative to most of the alternative models that have been proposed. These advantages include:

(1) Incomplete cases. In the layered model all available scores from each student are used, leading to less biased, more stable, more efficient estimates when compared to models which use gain scores as the response or which use prior scores as covariates.

(2) Adjustment for past scores. This is the *sine qua non* for a model to be considered a value-added model but different value-added models implement the adjustment in different ways. The layered model employs both within-student and between student adjustments which, particularly in the case when students fall into tracks (or strata), produces more defensible estimates.

(3) Adjustment for future scores. In models which jointly model multiple years of data, estimates can be "revised" as additional future data becomes available. In calculating the revised estimates, the layered model "rewards" teachers whose students go on to perform "unexpectedly well" in the future; the non-layered model (and low persistency general model) "punishes" those teachers.

(4) Omitted variables. The layered model is less susceptible to bias from omitted variables than is a non-layered model (or the low persistency general model).

Appendix 1. Technical Details, Part II: Estimates in Terms of "Cell Means"

For the type of data used in the examples in this paper, it is both pedagogically helpful and computationally more efficient to collapse the data from all students having the same teacher-sequence into a single "observation." How this is done is most easily seen by using Henderson's mixed model equations (Henderson, *et al.*, 1959; Searle, *et al.*, 1992, pp. 275-276).

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} \quad \text{so that} \quad \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} \quad \text{where}$$

b and **u** are the estimated fixed and random effects as before, and

$\mathbf{A}_{11} = \mathbf{X}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{A}_{12} = \mathbf{X}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Z}$, $\mathbf{A}_{12} = \mathbf{Z}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{X}$,

$\mathbf{A}_{22} = \mathbf{Z}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} = \mathbf{A}_{22(a)} + \mathbf{A}_{22(b)}$, and

$\mathbf{q}_1 = \mathbf{X}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}$, $\mathbf{q}_2 = \mathbf{Z}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}$.

Because **R** is a block-diagonal matrix with a block for each student, these matrices can be computed student-by-student, for example,

$$\mathbf{A}_{11} = \sum_{i} (\mathbf{X}_i^{T} \mathbf{R}_i^{-1} \mathbf{X}_i)$$

where the subscript "i" indicates a student. If the student has three years of data, then $\mathbf{X}_i$ has, three rows, one for each year, and $\mathbf{R}_i$ is 3-by-3. Thus $\mathbf{R}_i$ is much easier to invert than the full **R** matrix. Similar formulas hold for $\mathbf{A}_{12}$, $\mathbf{A}_{21}$, $\mathbf{A}_{22(a)}$, $\mathbf{q}_1$ and $\mathbf{q}_2$. Furthermore, all students in the same teacher sequence have the same value for $\mathbf{X}_i$, $\mathbf{Z}_i$, and $\mathbf{R}_i$, allowing further simplification, for example,

$$\mathbf{A}_{11} = \sum_{s} n_s (\mathbf{X}_s^{T} \mathbf{R}_s^{-1} \mathbf{X}_s) \quad \text{and} \quad \mathbf{q}_1 = \sum_{s} n_s (\mathbf{X}_s^{T} \mathbf{R}_s^{-1} \overline{\mathbf{y}}_s)$$

The subscript "s" indicates a teacher sequence, "n_s" is the number of students in that teacher sequence, and $\overline{\mathbf{y}}_s$ is the vector of mean scores (one mean score for each year) for students in that teacher sequence. Using this approach, it is possible to construct examples, and to solve problems involving real data, using teacher sequence averages rather than individual student scores (provided **R** and **G** are "known"). This is the method that was used for Example #1 in Tables 3 and 4.

Appendix 2. Technical Details, Part III: the Estimated Random Effects (u)

In section 2 it was noted that $\mathbf{u} = \mathbf{G}\mathbf{Z}^{T}\mathbf{V}^{-1}\mathbf{r}$. This formula can be shown to have a fairly "transparent" interpretation. Let **M** be a matrix that is partitioned (upper left & right, lower left & right) into $\mathbf{M}_{11}$, $\mathbf{M}_{12}$, $\mathbf{M}_{21}$, and $\mathbf{M}_{22}$. Let the corresponding submatrices of $\mathbf{M}^{-1}$ be $\mathbf{M}^{11}$, $\mathbf{M}^{12}$, $\mathbf{M}^{21}$, and $\mathbf{M}^{22}$. Finally, let $\mathbf{M}_{22.1} = \mathbf{M}_{22} - \mathbf{M}_{21}\,\mathbf{M}_{11}^{-1}\,\mathbf{M}_{12}$. Then it can be shown (Harville, 1997, p. 99) that

$$\mathbf{M}^{22} = [\mathbf{M}_{22.1}]^{-1};$$
$$\mathbf{M}^{12} = -\mathbf{M}_{11}^{-1}\,\mathbf{M}_{12}\,[\mathbf{M}_{22.1}]^{-1};$$
$$\mathbf{M}^{21} = -[\mathbf{M}_{22.1}]^{-1}\,\mathbf{M}_{21}\,\mathbf{M}_{11}^{-1};$$
$$\mathbf{M}^{11} = \mathbf{M}_{11}^{-1} + \mathbf{M}_{11}^{-1}\,\mathbf{M}_{12}\,[\mathbf{M}_{22.1}]^{-1}\,\mathbf{M}_{21}\,\mathbf{M}_{11}^{-1}.$$

Now let the n-by-n matrix $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^{T} + \mathbf{R}$ from the linear mixed model be partitioned so that the lower right submatrix, v_{22}, is a scalar, $\mathbf{V}_{11}$ is an (n−1)-by-(n−1) matrix, and $\mathbf{v}_{12}$ is an (n−1)-by-1 vector. Applying the above matrix identities produces the following useful results concerning the elements of $\mathbf{V}^{-1}$:

$$v^{22} = 1 / (v_{22} - \mathbf{v}_{12}^{T}\,\mathbf{V}_{11}^{-1}\,\mathbf{v}_{12}) = 1 / \sigma^2_{2.1}\ ;$$
$$\mathbf{v}^{12} = -\mathbf{V}_{11}^{-1}\,\mathbf{v}_{12} / \sigma^2_{2.1} = -\boldsymbol{\beta}_{2.1} / \sigma^2_{2.1}\ .$$

$\boldsymbol{\beta}_{2.1}$ contains the regression coefficients, and $\sigma^2_{2.1}$ is the conditional ("error") variance, from the regression of the last element of **r** (the n-th element, r_n) on $\mathbf{r}_{not\text{-}n}$ which contains the other (n−1) elements of **r**. In general, then, in any row/column of $\mathbf{V}^{-1}$ (say the i-th row/column), the diagonal element contains $1 / \sigma^2_{i.(not\text{-}i)}$ and the off-diagonal elements contain $-\boldsymbol{\beta}_{i.(not\text{-}i)} / \sigma^2_{i.(not\text{-}i)}$. Consequently, $\mathbf{V}^{-1}\mathbf{r}$ is an n-by-1 vector in which the i-th element is

$$(r_i - \mathbf{r}_{not\text{-}i}^{T}\,\boldsymbol{\beta}_{i.(not\text{-}i)}) / \sigma^2_{i.(not\text{-}i)} = d_{i.(not\text{-}i)} / \sigma^2_{i.(not\text{-}i)}.$$

The numerator of this quantity ($d_{i.(not-i)}$) is just a "regression residual": the difference between an "observed deviation score" and an "expected deviation score" whose expectation is based on the information in $\mathbf{r}_{not-i}$. Let **d** be the vector of all the $d_{i.(not-i)}$ values, and let **S** be a diagonal matrix of $\sigma^2_{i.(not-i)}$ values. Then

$$\mathbf{V}^{-1}\mathbf{r} = \mathbf{S}^{-1}\mathbf{d}.$$

The formula for any particular element of **u**, the j-th element for example, is

$$u_j = g_{jj}\, \mathbf{z}_j^T\, \mathbf{V}^{-1}\, \mathbf{r} = g_{jj}\, \mathbf{z}_j^T\, \mathbf{S}^{-1}\, \mathbf{d} = (1 / n_j) \cdot \mathbf{z}_j^T \cdot \text{diag}[n_j\, (g_{jj} / (\sigma^2_{i.(not-i)})] \cdot \mathbf{d}$$

where g_{jj} is the j-th diagonal element of **G**, $\mathbf{z}_j$ is the jth column of **Z**, n_j is the sum of the elements of $\mathbf{z}_j$, and "diag[…]" is a diagonal matrix. In the layered teacher model, the elements of $\mathbf{z}_j$ are (under most circumstances) zeros and ones so that n_j is simply the number of non-zero elements in $\mathbf{z}_j$. The estimated "effect" for a particular teacher (u_j) is then calculated as follows: **d** is a vector of observed minus expected deviation scores; diag[n_j (g_{jj} / ($\sigma^2_{i.(not-i)}$)] "shrinks" the elements of **d** toward zero; $\mathbf{z}_j^T$ adds up the shrunken **d** values corresponding to the non-zero elements of $\mathbf{z}_j$; dividing by n_j gets the average.

References

Ballou, D., Sanders, W. L., and Wright, S. P. (2004). Controlling for Student Background in Value-Added Assessment of Teachers. *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, pp. 37-65.

Collins, L. M., Schafer, J. L., and Kam, C.-M. (2001). A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures. *Psychological Methods*, Vol. 6, No. 4, pp. 330-351.

Doran, H. C. (2003). Adding Value to Accountability. *Educational Leadership*, Vol. 61, No. 3, pp. 55-59. Available at www.cgp.upenn.edu/ope_news.html.

Drury, D. and Doran, H. C. (2003). The Value of Value-Added Analysis. *Policy Research Brief for National School Board Association*, Vol. 3, No. 1. Available at www.cgp.upenn.edu/ope_news.html.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag.

Henderson, C. R., Kempthorne, O., Searle, S. R., and Von Krosig, C. N. (1959). Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics*, Vol. 30, pp. 583-588.

Hershberg, T., Simon, V. A., and Lea-Kruger, B. (2004). Measuring What Matters. *American School Board Journal*, Vol. 191, No. 2. Available at www.cgp.upenn.edu/ope_news.html.

Kiplinger, V. L. (2004). Longitudinal Study of Student Growth in Reading, Writing and Mathematics Achievement in Thirty-six Colorado Public School Districts, Phase II, Final Report. Available at http://matrix10.asd20.org/lssg/phase2/Phase2Report.pdf.

Little, R. J. A., and Rubin, D. B. (2002). *Statistical Analysis with Missing Data, Second Edition*. New York: Wiley.

Lissitz, R. W., editor (2005). *Value Added Models in Education: Theory and Applications.* Maple Grove, MN: JAM Press.

McCaffrey, D. F., Lockwood, J. R., Mariano, L. T., and Setodji, C. (2005). Challenges for Value-Added Assessment of Teacher Effects. In Lissitz (Ed.), *Value Added Models in Education: Theory and Applications.*, Chapter 5, pp. 111-144.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., and Hamilton, L. (2004). Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, pp. 67-101.

Raudenbush, R. W. (2004). What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice? *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, pp. 121-129.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousands Oaks, CA: Sage Publications.

Robinson, G. K. (1991). That BLUP Is a Good Thing: The Estimation of Random Effects. *Statistical Science*, Vol. 6, No. 1, pp. 15-51.

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, Vol. 100, No. 469, pp.322-.331.

Rubin, D. B., Stuart, E. A., and Zanutto, E. L. (2004). A Potential Outcomes View of Value-Added Assessment in Education. *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, pp. 103-116.

Sanders, W. L., Saxton, A. M., and Horn, S. P. (1997). The Tennessee Value-Added Assessment System: A Quantitative Outcomes-Based Approach to Educational Assessment. In J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Educational Measure?* pp. 137-162. Thousand Oaks, CA: Corwin Press.

Searle, S.R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. New York: John Wiley and Sons.

Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., and Resnick, M. B. (2004). An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance. *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, pp. 11-36.

Wainer, H., editor (2004). Value-Added Assessment Special Issue. *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1.

Webster, W. J. (2005). The Dallas School-Level Accountability Model: The Marriage of Status and Value-Added Approaches. In Lissitz (Ed.), *Value Added Models in Education: Theory and Applications.*, Chapter 9, pp. 233-271.

Urquhart, N. S. (1982). Adjustment in Covariance when One Factor Affects the Covariate. *Biometrics*, Vol. 38, pp.651-660.

Table 1. Example #1 with No Shrinkage

ID	Teacher	Score	X	Z			Q_1	Dev	Q_2			u	Est. Mean
01	A	375	1	1	0	0	1/3	−25	1	0	0	−25	375
02	B	400	1	0	1	0	1/3	0	0	1	0	0	400
03	C	425	1	0	0	1	1/3	+25	0	0	1	+25	425

Table 2. Example #1 with Shrinkage

ID	Teacher	Score	Q_1	Dev	Q_2			u	Est. Mean
01	A	375	1/3	−25	0.5	0	0	−12.5	387.5
02	B	400	1/3	0	0	0.5	0	0	400
03	C	425	1/3	+25	0	0	0.5	+12.5	412.5

Table 3. Example #1 with Unequal Ns and No Shrinkage

Teacher	Mean Score	N	Q_1	Mean Dev	Q_2			u	Est. Mean
A	375	1	1/3	−25	1	0	0	−25	375
B	400	2	1/3	0	0	1	0	0	400
C	425	5	1/3	+25	0	0	1	+25	425
All	412.5	8							400

Table 4. Example #1 with Unequal Ns and Shrinkage

Teacher	Mean Score	N	Q_1	Mean Dev	Q_2			u	Est. Mean
A	375	1	1/4	−29.17	1/2	0	0	−14.58	389.58
B	400	2	1/3	−4.17	0	2/3	0	−2.78	401.39
C	425	5	5/12	+20.83	0	0	5/6	+17.36	421.53
All	412.5	8							404.17

Table 5. Example #2

						Z (NLM)				Z (LM)			
ID	**Year**	**Teacher**	**Score**	**X1**	**X2**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
01	1	A	275	1	0	1	0	0	0	1	0	0	0
	2	C	355	0	1	0	0	1	0	1	0	1	0
02	1	A	285	1	0	1	0	0	0	1	0	0	0
	2	D	405	0	1	0	0	0	1	1	0	0	1
03	1	B	325	1	0	0	1	0	0	0	1	0	0
	2	C	405	0	1	0	0	1	0	0	1	1	0
04	1	B	315	1	0	0	1	0	0	0	1	0	0
	2	D	435	0	1	0	0	0	1	0	1	0	1

Table 6. Estimates for Example #2

	Year	**Teacher**	**NLM(0)**	**NLM(0.7)**	**LM(0)**	**LM(0.7)**
b	1		300	300	300	300
	2		400	400	400	400
u	1	A	−20	−6	−20	−20
		B	+20	+6	+20	+20
	2	C	−20	−20	−20	−20
		D	+20	+20	+20	+20

Table 7. Q_2 for Teacher A in Example #2

ID	Year	Teacher	Score	Deviation	NLM(0)	NLM(0.7)	LM(0)	LM(0.7)
1	1	A	275	−25	0.5	0.5	0.375	0.375
	2	C	355	−45	0	−0.175	0.125	0.125
2	1	A	285	−15	0.5	0.5	0.375	0.375
	2	D	405	+5	0	−0.175	0.125	0.125
3	1	B	325	+25	0	0	0.125	0.125
	2	C	405	+5	0	0.175	−0.125	−0.125
4	1	B	315	+15	0	0	0.125	0.125
	2	D	435	+35	0	0.175	−0.125	−0.125

Table 8. Q_2 for Teacher C in Example #2

ID	Year	Teacher	Score	Deviation	NLM(0)	NLM(0.7)	LM(0)	LM(0.7)
1	1	A	275	−25	0	−0.175	−0.25	−0.425
	2	C	355	−45	0.5	0.5	0.5	0.5
2	1	A	285	−15	0	0.175	−0.25	−0.075
	2	D	405	+5	0	0	0	0
3	1	B	325	+25	0	−0.175	−0.25	−0.425
	2	C	405	+5	0.5	0.5	0.5	0.5
4	1	B	315	+15	0	0.175	−0.25	−0.075
	2	D	435	+35	0	0	0	0

Table 9. Formulas for Example #2

NLM(0)	**A**	$M1_A$.
	C	$M2_C$.
NLM(0.7)	**A**	$M1_A - 0.35\ M2_A + 0.35\ M2_B\ =\ M1_A - 0.70\ (M2_A - M2_{AB})$.
	C	$M2_C - 0.35\ M1_C + 0.35\ M1_D\ =\ M2_C - 0.7\ (M1_C - M1_{CD})$.
LM(0)	**A**	$0.75\ M1_A + 0.25\ M1_B + 0.25\ M2_A - 0.25\ M2_B$ $=\ M1_A + (1/2)[(M2_A - M2_{AB}) - (M1_A - M1_{AB})]$.
	C	$M2_C - M1_{AB}$.
LM(0.7)	**A**	Same as LM(0)
	C	$M2_C - 0.85\ M1_C - 0.15\ M1_D$ $=\ M2_C - 0.5\ M1_C - 0.5\ M1_D - 0.7\ M1_C + 0.35\ M1_C + 0.35\ M1_D$ $=\ M2_C - M1_{AB} - 0.7\ (M1_C - M1_{CD})$.

Table 10. Example #3

ID	Year	Teacher	Score	Q_1: Yr1	Q_1: Yr2	Q_1: Yr3
01	1	A	370	1/12	0	0
	2	D	440	0	1/12	0
	3	G	470	0	0	1/12
02	1	A	350	1/12	0	0
	2	D	380	0	1/12	0
	3	H	530	0	0	1/12
03	1	A	330	1/12	0	0
	2	E	500	0	1/12	0
	3	G	550	0	0	1/12
04	1	A	310	1/12	0	0
	2	E	440	0	1/12	0
	3	H	610	0	0	1/12
05	1	B	430	1/12	0	0
	2	D	500	0	1/12	0
	3	G	530	0	0	1/12
06	1	B	410	1/12	0	0
	2	D	440	0	1/12	0
	3	H	590	0	0	1/12
07	1	B	390	1/12	0	0
	2	E	560	0	1/12	0
	3	G	610	0	0	1/12
08	1	B	370	1/12	0	0
	2	E	500	0	1/12	0
	3	H	670	0	0	1/12
09	1	C	460	1/3	0	0
	2	F	560	0	1/3	0
	3	J	660	0	0	1/3

Table 11. Estimates for Example #3

	Year	Teacher	NLM(0)	NLM(0.7)	LM(0)	LM(0.7)	Mean Score	Mean Gain
b	1		400	400	400	400	380	
	2		500	500	500	500	480	100
	3		600	600	600	600	580	100
u	1	A	−60	−35.3	−60	−60	340	
		B	0	−24.7	0	0	400	
		C	+60	+60	+60	+60	460	
	2	D	−60	−51.8	−35	−49	440	50
		E	0	−8.2	+35	+49	500	150
		F	+60	+60	0	0	560	100
	3	G	−60	−76.5	−30	−46.5	540	40
		H	0	+16.5	+30	+46.5	600	160
		J	+60	+60	0	0	660	100

Table 12. Q_2 for Year-2 Teachers D and F in Example #3

				NLM(0)		NLM(0.7)		LM(0)		LM(0.7)	
ID	**Year**	**Tchr**	**Dev**	**D**	**F**	**D**	**F**	**D**	**F**	**D**	**F**
01	1	A	−30	0	0	−.0515	0	−.125	0	−.2125	0
	2	D	−60	0.25	0	+.25	0	+.1875	0	+.1875	0
	3	G	−130	0	0	−.0515	0	+.0625	0	+.0625	0
02	1	A	−50	0	0	−.0515	0	−.125	0	−.2125	0
	2	D	−120	0.25	0	+.25	0	+.1875	0	+.1875	0
	3	H	−70	0	0	−.0515	0	+.0625	0	+.0625	0
03	1	A	−70	0	0	+.0515	0	−.125	0	−.0375	0
	2	E	0	0	0	0	0	+.0625	0	+.0625	0
	3	G	−50	0	0	+.0515	0	−.0625	0	−.0625	0
04	1	A	−90	0	0	+.0515	0	−.125	0	−.0375	0
	2	E	−60	0	0	0	0	+.0625	0	+.0625	0
	3	H	+10	0	0	+.0515	0	−.0625	0	−.0625	0
05	1	B	+30	0	0	−.0515	0	−.125	0	−.2125	0
	2	D	0	0.25	0	+.25	0	+.1875	0	+.1875	0
	3	G	−70	0	0	−.0515	0	+.0625	0	+.0625	0
06	1	B	+10	0	0	−.0515	0	−.125	0	−.2125	0
	2	D	−60	0.25	0	+.25	0	+.1875	0	+.1875	0
	3	H	−10	0	0	−.0515	0	+.0625	0	+.0625	0
07	1	B	−10	0	0	+.0515	0	−.125	0	−.0375	0
	2	E	+60	0	0	0	0	+.0625	0	+.0625	0
	3	G	+10	0	0	+.0515	0	−.0625	0	−.0625	0
08	1	B	−30	0	0	+.0515	0	−.125	0	−.0375	0
	2	E	0	0	0	0	0	+.0625	0	+.0625	0
	3	H	+70	0	0	+.0515	0	−.0625	0	−.0625	0
09	1	C	+60	0	0	0	0	0	−1	0	−1
	2	F	+60	0	+1	0	+1	0	+1	0	+1
	3	J	+60	0	0	0	0	0	0	0	0

Table 13. Q_2 for Year-3 Teachers G and J in Example #3

				NLM(0)		NLM(0.7)		LM(0)		LM(0.7)	
ID	**Year**	**Tchr**	**Dev**	**G**	**J**	**G**	**J**	**G**	**J**	**G**	**J**
01	1	A	−30	0	0	−.0515	0	0	0	−.0515	0
	2	D	−60	0	0	−.0515	0	−.125	0	−.1765	0
	3	G	−130	0.25	0	+.25	0	+.25	0	+.25	0
02	1	A	−50	0	0	+.0515	0	0	0	+.0515	0
	2	D	−120	0	0	+.0515	0	−.125	0	−.0735	0
	3	H	−70	0	0	0	0	0	0	0	0
03	1	A	−70	0	0	−.0515	0	0	0	−.0515	0
	2	E	0	0	0	−.0515	0	−.125	0	−.1765	0
	3	G	−50	0.25	0	+.25	0	+.25	0	+.25	0
04	1	A	−90	0	0	+.0515	0	0	0	+.0515	0
	2	E	−60	0	0	+.0515	0	−.125	0	−.0735	0
	3	H	+10	0	0	0	0	0	0	0	0
05	1	B	+30	0	0	−.0515	0	0	0	−.0515	0
	2	D	0	0	0	−.0515	0	−.125	0	−.1765	0
	3	G	−70	0.25	0	+.25	0	+.25	0	+.25	0
06	1	B	+10	0	0	+.0515	0	0	0	+.0515	0
	2	D	−60	0	0	+.0515	0	−.125	0	−.0735	0
	3	H	−10	0	0	0	0	0	0	0	0
07	1	B	−10	0	0	−.0515	0	0	0	−.0515	0
	2	E	+60	0	0	−.0515	0	−.125	0	−.1765	0
	3	G	+10	0.25	0	+.25	0	+.25	0	+.25	0
08	1	B	−30	0	0	+.0515	0	0	0	+.0515	0
	2	E	0	0	0	+.0515	0	−.125	0	−.0735	0
	3	H	+70	0	0	0	0	0	0	0	0
09	1	C	+60	0	0	0	0	0	0	0	0
	2	F	+60	0	0	0	0	0	−1	0	−1
	3	J	+60	0	+1	0	+1	0	+1	0	+1

Table 14. Formulas for Example #3

NLM(0)	**D**	$M2_D$
	G	$M3_G$
NLM(0.7)	**D**	$M2_D - (b/2)\ M1_D + (b/2)\ M1_E - (b/2)\ M3_D + (b/2)\ M3_E$ $= M2_D - b\ (M1_D - M1_{DE}) - b\ (M3_D - M3_{DE})$.
	G	$M3_G - (b/2)\ M2_G + (b/2)\ M2_H - (b/2)\ M1_G + (b/2)\ M1_H$ $= M3_G - b\ (M2_G - M2_{GH}) - b\ (M1_G - M1_{GH})$.
LM(0)	**D**	$0.75\ M2_D + 0.25\ M2_E - 0.5\ M1_D - 0.5\ M1_E + 0.25\ M3_D - 0.25\ M3_E$ $= M2_D - 0.25\ (M2_D - M2_E) + 0.25\ (M3_D - M3_E) - M1_{DE}$ $= M2_D - M1_{DE} + .5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})]$.
	G	$M3_G - (1/2)\ M2_G - (1/2)\ M2_H = M3_G - M2_{GH}$.
LM(0.7)	**D**	$0.75\ M2_D + 0.25\ M2_E + 0.25\ M3_D - 0.25\ M3_E - 0.85\ M1_D - 0.15\ M1_E$ $= M2_D - .25\ (M2_D - M2_E) + .25\ (M3_D - M3_E)$ $- 0.5\ M1_D - 0.5\ M1_E - 0.35\ M1_D + 0.35\ M1_E$ $= M2_D - M1_{DE} - 0.7\ (M1_D - M1_{DE})$ $+ .5[(M3_D - M3_{DE}) - (M2_D - M2_{DE})]$.
	G	$M3_G - [(1+b) / 2]\ M2_G - [(1-b) / 2]\ M2_H - (b/2)\ M1_G + (b/2)\ M1_H$ $= M3_G - M2_{GH} - b\ (M2_G - M2_{GH}) - b\ (M1_G - M1_{GH})$.