Cross-Classified Models in the Context of Value-Added Modeling *

J. Kyle Roberts †
Southern Methodist University

Douglas Bates
University of Wisconsin- Madison

1 Background to Value-Added Modeling

In the past few years, value-added modeling (VAM) has become increasingly popular as a means for monitoring student knowledge growth and teacher accountability. Initially, VAMs focused on model development (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Tekwe et al., 2004) and the development of software routines appropriate for estimating these models. Presently, researchers are beginning to question other aspects of VAMs, such as whether these estimates can reasonably be thought of as causal (Rubin, Stuart, & Zanutto, 2004) and the appropriateness of cross-classified models for teacher accountability (Lockwood, Doran & McCaffrey, 2003).

Work by Doran, Bates, and Phillips (under review) and work presented in the special issue of the Journal of Educational and Behavioral Statistics, volume 29(1) have made considerable strides above and beyond initial VAMs. Specifically, Doran et al. are building models which honor the psychometric properties of the instruments used in VAMs. Furthermore, their findings indicate that the VAM school effects are not invariant to the choice of an item response theory model.

If models indeed are going to be built for teachers (and schools and districts), a cross-classified model (Lockwood, Doran & McCaffrey, 2003; Raudenbush & Bryk, 2002) is preferred over simpler multilevel models. This class of statistics models the property that teacher (or school and district) effects have an additive characteristic to student continual performance, such that past teachers may affect future outcomes. The difficulty is that these types of models are complex to build and computationally demanding. Recent developments in statistical software routines have also paved the way for the development of more complex VAMs. Developments in packages like `lme4` in the R software package have made these more complex models easily estimable through advances in the sparse Cholesky decomposition (Bates, 2008). The strength with handling data in this type of way is that prior and future effects of teachers/schools/districts can be modeled as random effects and their potential value to a student's growth towards adequate yearly progress can be correctly modeled. Not

*Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, March 6, 2010, Washington, D.C.

†The first author may be reached at kyler@smu.edu

honoring this model structure could inadvertently inflate (or deflate) the perceived value-added for a given teacher if students coming in to the teacher's class previously had highly effective (or less effective) teachers.

In this paper, we aim to evaluate current models of VAM and investigate the effectiveness of each model in terms of their adequacy for monitoring student adequate yearly progress, teacher effectiveness, school effectiveness, and district accountability. We also discuss the continued development of the `lme4` package in R to handle complex cross-classified models for teacher accountability with exceptionally large datasets.

1.1 Current Models for VAMs

One of the first true VAMs was introduced by Sanders, Saxton & Horn (1997). In this work, the Tennessee Value-Added Assessment System was first articulated, but the authors quickly noted that a model like this had tremendous computational requirements. Although great strides have been made in the area of software development, Lockwood et al. (2007) has recently noted that there are still computational problems for solving VAMs for large datasets and hence large districts and states.

VAMs typically fall into one of two categories: models appropriate for monitoring school effects on student learning and models appropriate for monitoring teacher effects on student learning. These models typically take the form:

$$Y_{ti} = [\beta_0 + \beta_1(year_t)] + \left[\theta_{0j(i)} + \theta_{1j(i)}(year_t) + \delta_{0i} + \delta_{1i}(year_t) + \epsilon_{ti}\right] \tag{1}$$

where i indexes students and j could index either the schools or teachers. The notation $j(i)$ represents the school (or teacher) attended by student i in year t. In this model, the within-group errors are assumed IID where $\epsilon_{ti} \sim \mathcal{N}(0, \sigma^2)$ and school (or teacher) and student random effects are $\boldsymbol{\theta}_j = (\theta_{0j}, \theta_{1j}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ and $\boldsymbol{\delta}_i = (\delta_{0i}, \delta_{1i}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, respectively.

Furthermore, covariates may be added as fixed effects to Equation 1 such that:

$$Y_{ti} = [\beta_0 + \beta_1(year_t) + \beta_q x_{qti}] + \left[\theta_{0j(i)} + \theta_{1j(i)}(year_t) + \delta_{0i} + \delta_{1i}(year_t) + \epsilon_{ti}\right] \tag{2}$$

where x_{qti} represents the vector of all student-level covariates for student i at time point t. In this case, all of the covariates are modeled as fixed effects.

There is a subtle, but important, oversight in our notation in equations (1) and (2) that illustrates the computational challenges in fitting VAMs to large longitudinal (i.e. following students over time) data sets. We write $j(i)$ for the school (or teacher's class) attended by student i at time t whereas it should be written $j(t, i)$. The data sets described in section 2.3 include results from several years of each student's career during which time most students will change schools and nearly all students will change teachers. So if j is the teacher index then the expression $j(i)$ is not well-defined because student i has different teachers in different years.

The methods for fitting hierarchical linear models (Raudenbush & Bryk, 2002), also called multilevel models (Goldstein, 1995), that are well-known to education researchers, cannot be applied to such data. They only apply to data where the hierarchical classification of the scores by, say, students within classes (teachers) within schools within districts, etc. are constant over time. To obtain reliable estimates of learning growth by students and the contributions associated with the teachers and schools to which they have been exposed we must follow the students' progress for several years with different teachers and often in different schools. When we observe the same student with different teachers (or in different schools) the student and teacher effects (student and school effects) are no longer nested; they become partially "crossed". Methods for fitting value-added models to such data must allow for cross-classified random effects.

It is sometimes the convention for districts and states using VAMs to employ a two-stage model whereby initial covariates are introduced into an Ordinary Least Squares (OLS) analysis and then run subsequent mixed-effects models on student achievement scores (e.g., Maryland High School Graduation Exam, Dallas ISD, and Texas Education Agency). It makes little sense to include variables into an OLS analysis and then suppose that these variables will have the same type of effect in a multilevel analysis. In a discussion of the procedures for this type of model, the Texas Education Agency argues for this type of two-stage model arguing that "All predictor variables [from the OLS] were significant at an alpha level of 0.05. Thus, all initial predictors were eligible to be included in a multilevel model." (TEA, 2009; p. 35). This methodology makes little sense considering that multilevel analysis has as its basis the thought that measurement units are NOT independent from the context in which they are drawn and OLS has as an assumption the idea that measurement units are independent from the context in which they are drawn. Put another way, OLS assumes that there is no between-school differences and multilevel analysis tries to model the fact that there is between-school differences (cf. Goldstein, 1995; Hox, 2002; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Furthermore, any covariate that a VAM may wish to model could easily be incorporated into the multilevel model without the need for this type of two-stage convention.

However, it is presently not known if the addition of certain covariates to a VAM are warranted. Sanders and Wright (2008) have argued that SES adjustments to VAMs are unneeded when the number of measurement occasions increases. They also argue that since teacher assignment patterns are typically related to teacher effectiveness, the effect of modeling these SES covariates could over-adjust student scores.

When building any type of growth model, the number of time points included in the model is of prime concern if indeed the goal of the VAM is to identify the building of a trend (Sanders & Wright, 2008). Growth models that are based solely on two time points are notoriously circumspect as to their accuracy since there is an increased probability that ability estimates for a student will be affected to a negative degree. For example, suppose that within one of the two years of classification, a student does exceptionally poorly on an exam for no other reason than they had a bad day. This would negatively affect the student's projections and introduces some degree of error into the model. This is further exacerbated when only two measurement occasions are used to estimate student growth. In simple OLS regression, the consequences of such an action mean that the slope for a regression line is attenuated (biased) towards zero. The consequences of this action in a multilevel framework could mean that schools with a disproportionate number of students entering the school who achieve at high levels will appear "better" than they actually are, and conversely schools in which a large number of students are entering the school with lower scores could make the school appear "worse." Any model that might potentially be used to measure AYP for school districts, campuses, or classrooms that does not consider more than two time points is needlessly introducing measurement error into the model. The addition of more time points will not only help the precision of the model, but it will also reduce the number of false positive and false negative district (/school/teacher) effects.

McCaffrey, Lockwood, Koretz & Hamilton (2003) raised some serious issues as to the usability of the results from VAMs. Among the questions introduced in their research are the problems concerning omitted variables and missing data, problems concerning the type and nature of the instruments used to measure student achievement, problems with modeling measurement error, and problems with using achievement measures as proxies for estimating teacher effectiveness. As part of their research, McCaffrey et al. (2003) made seven recommendations for future research:

1. Develop databases that can support VAM estimation of teacher effects.
2. Develop computational tools for fitting VAM
3. Link VAM teacher-effect estimates to alternative measures of teacher effectiveness
4. Empirically evaluate the potential sources of errors we have identified
5. Estimate the prevalence of factors that contribute to the sensitivity of teacher-effect estimates
6. Incorporate decision theory into VAM
7. Use research and auxiliary data to inform modeling choices (pp. 114-119)

Several researchers have already begun working to contribute to answering some of these calls for research. For example, Doran, Bates, and Phillips (under review) have turned their attention to investigating the consequences of psychometric decisions on VAMs for school effects (Recommendation 4).

1.2 The Strength of Cross-Classified VAMs

Cross-classified models are a special case of mixed-effects models in which the typical assumption of complete nesting in hierarchical linear modeling is not met. For data to be considered purely clustered, clusters of elements from one level must all belong to single elements of a higher level. For example, in the following table, students from five separate high schools are completely nested inside each of the five high schools.

	School 1	School 2	School 3	School 4	School 5
Student	1,2,3	4,5	6,7,8	9,10	11,12,13,14

In the above example, each student is nested inside one and only one school. This data may be modeled in a mixed-effects model as:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \tag{3}$$

where γ_{00} represents the average student achievement across all j schools with student-level residuals e_{ij} and school-level random effects u_{0j}. This type of conventional modeling is sometimes referred to as hierarchical linear modeling (Raudenbush & Bryk, 2002), multilevel modeling (Hox, 2002), and mixed-effects modeling (Pinheiro & Bates, 2000).

However, consider an effect where student achievement scores are modeled as a product of both which middle school they attend and which high school they attend. In an ideal setting, students from each middle school would matriculate into one and only one high school. In this case, the data might be represented as:

	Middle School					
High School	MS 1	MS 2	MS 3	MS 4	MS 5	MS 6
HS 1	S1, S2	S3, S4, S5				
HS 2			S6, S7	S8, S9, S10		
HS 3					S11, S12, S13	S14, S15

In this example, each student is nested inside one and only one middle school and these middle schools are nested inside three high schools. This data may be modeled in a mixed-effects model as:

$$Y_{ijk} = \gamma_{000} + u_{00k} + u_{0jk} + e_{ijk} \tag{4}$$

where γ_{000} represents the average student achievement across all j middle schools and k high schools with student-level residuals e_{ijk} and middle school-level random effects u_{0jk} and high school-level random effects u_{00k}.

Although the above representation is an ideal setting, it is more the convention that students are actually cross-classified by both middle school and high school such that their data structure resemble:

	Middle School					
High School	MS 1	MS 2	MS 3	MS 4	MS 5	MS 6
HS 1	S1, S2	S3	S4		S5	
HS 2		S6, S7		S8, S9, S10		
HS 3	S11		S12		S13	S14, S15

In this cross-classified example, student scores are modeled as:

$$Y_{i(j_1,j_2)} = \gamma_{000} + u_{00j_1} + u_{00j_2} + u_{00j_1xj_2} + e_{i(j_1,j_2)} \tag{5}$$

where γ_{000} represents the average student achievement for student i having attended middle school j_1 and high school j_2 with random effects u_{00j_1} for the middle schools, u_{00j_2} for the high schools, a random interaction effect $u_{00j_1xj_2}$ for belonging to both middle school j_1 and high school j_2 and student-level residuals $e_{i(j_1,j_2)}$. This parameterization follows the Rasbash and Browne (2001) notation.

This type of cross-classified modeling has strong applications in monitoring teacher and school effects for student achievement scores across time. In the case where the object of modeling is teacher effectiveness, the cross-classified model is the only convention for modeling time series data unless the measurement occasions are limited to just one year. Consider just the simple case where two years worth of student data (7th and 8th grades) are modeled in a cross-classified setting. In this case, our model would take on the form:

$$Y_{ti(j_1,j_2)} = \gamma_{0000} + u_{000j_1} + u_{000j_2} + u_{000j_1xj_2} + u_{00ij_1} + u_{00ij_2} + u_{00ij_1xj_2} + e_{ti(j_1,j_2)}. \tag{6}$$

As can be seen from the above equation, these models become exponentially complex the more years worth of data and the more grades we add to a model. Indeed, Lockwood et al. (2007) note that these models can become so increasingly complex that more efficient algorithms need to be developed in order to conduct further research into cross-classified models. They also propose a notation slightly different from the Rasbash and Browne (2001) notation where teacher effects are modeled as:

$$Y_{ti} = \mu_t + \beta_t^{'} x_{it} + \sum_{t^* \leq t} \alpha_{tt^*} \phi_{it^*}^{'} \theta_{t^*} + \epsilon_{it} \tag{7}$$

where μ_t is the overall mean for each year, x_{it} is the covariate vector of both time invariant and time varying variables, θ_t are the teacher effects for each year, ϕ_{it} are the student effects vectors, and ϵ_{ti} is the residual score for student i at time t.

2 Challenges for Current VAMs

2.1 Computational Challenges for Software Packages and the Strength of `lme4`

To obtain reliable estimates of the value-added to student achievement scores by teachers and/or schools we will ideally have several scores for each student over time and information on the teacher and/or school at each of the measurement occasions. Unfortunately, the data attributes – large longitudinally-linked databases providing achievement results for several years classified by student and teacher and/or school – that are most beneficial in estimating VAMs also create the greatest computational challenges in fitting these models.

Because describing the computational methods using the scalar notation, such as $Y_{t,i}$, for individual scores quickly becomes unmanageable, we will switch to a vector/matrix notation. We write the entire set of responses for all students across each of the testing occasions for each student as $\boldsymbol{y}$. If there are 15,000 students and an average of 20 measurements per student then $\boldsymbol{y}$ is a vector of length 300,000. This may seem large but it is not large by current standards. Many state-wide testing programs now have databases of several millions of observations on hundreds of thousands of students, tens of thousands of teachers and thousands of schools. Any practical system for fitting cross-classified VAM models must allow for very large databases.

Let n be the total number of observations being modeled. The fixed-effects covariates, which are usually demographic characteristics such as sex, race and socioeconomic status, are used to create an $n \times p$ model matrix $\boldsymbol{X}$. Each column of $\boldsymbol{X}$ is associated with a particular coefficient in the fixed-effects part of the model. For example, a coefficient could represent the difference in performance of male and female students. We write the vector of random-effects coefficients as $\boldsymbol{\beta}$. The description "fixed" refers to the fact that the number of columns in $\boldsymbol{X}$ and the interpretations of the particular columns are fixed. If we obtain data from the next set of tests and add it to the database then re-fit the model, we will change neither the number of columns, p, nor the interpretation of the coefficients. Typically p is rather small. For example, there may be more than 10 columns in $\boldsymbol{X}$ but usually not more than 100.

The random-effects model matrices, on the other hand, usually have a very large number of columns and the number of columns can grow as we incorporate more data into the model. Suppose we use random effects for students, teachers and schools in a model for data with a total of n_1 students, n_2 teachers and n_3 schools. If we allow for two random effects for students (the random effect for the intercept and for the slope with respect to time — δ_{0i} and δ_{1i} in equations (1) and (2)) then the model matrix $\boldsymbol{Z}_1$ for the student random effects will be $n \times 2n_1$. The random-effects model matrices $\boldsymbol{Z}_2$ for teachers and $\boldsymbol{Z}_3$ for schools would have dimension $n \times 2n_2$ and $n \times 2n_3$, respectively. Let $\boldsymbol{b}$, a vector of length $q = 2(n_1 + n_2 + n_3)$, be the vector of all random effects (students, teachers and schools) and $\boldsymbol{Z} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3]$ be the corresponding model matrix.

The value-added model could then be written

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n),\ \boldsymbol{b} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right) \tag{8}$$

where $\boldsymbol{\theta}$ is the vector of variance components to be estimated. For the model we are describing $\boldsymbol{\theta}$ would have a total of nine elements; two variances and one covariance for each of the student, teacher and school classifications.

Because the matrix $\boldsymbol{Z}$ can be huge (perhaps millions of rows and hundreds of thousands of columns) this representation would be interesting but completely impractical if it were not for the fact that $\boldsymbol{Z}$ is very, very sparse. In the example we are describing each row is zero in all but six of the columns (two student random effects, two teacher random effects and two school random effects). Such matrices can be stored and manipulated using sparse matrix techniques (Davis, 2006), even for databases of the sizes we are discussing.

One of the shining successes of sparse matrix theory and software is the ability to calculate the Cholesky decomposition of large, sparse, symmetric, positive-definite matrices. Fortuitously, the calculations needed to evaluate the log-likelihood for the parameters in a model like (8) can be reduced to exactly such a calculation. Without going into details, the broad outlines are to express the variance-covariance of the random effects, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, in terms of a relative covariance factor, $\boldsymbol{\Lambda}(\boldsymbol{\theta})$, defined so that

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2\Lambda(\boldsymbol{\theta})\boldsymbol{\Lambda}'(\boldsymbol{\theta}) \tag{9}$$

and optimize the profiled deviance, defined as

$$-2\tilde{\ell}(\boldsymbol{\theta}) = \log(|\boldsymbol{L}(\boldsymbol{\theta})|^2) + n\left[1 + \frac{2\pi r^2(\boldsymbol{\theta})}{n}\right] \tag{10}$$

where $\boldsymbol{L}(\boldsymbol{\theta})$ is the sparse, lower-triangular Cholesky factor satisfying

$$\boldsymbol{L}(\boldsymbol{\theta})\boldsymbol{L}'(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\boldsymbol{\theta})'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Lambda}(\boldsymbol{\theta}) + \boldsymbol{I}. \tag{11}$$

In (10) $|\boldsymbol{L}(\boldsymbol{\theta})|$ denotes the determinant of $\boldsymbol{L}$, which is simply the product of its diagonal elements, and $r^2(\boldsymbol{\theta})$ is the minimum penalized residual sum of squares, which can be written as

$$r^2(\boldsymbol{\theta}) = \min_{\boldsymbol{u},\boldsymbol{\beta}} \left\| \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{Z\Lambda}(\boldsymbol{\theta}) & \boldsymbol{X} \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\beta} \end{bmatrix} \right\|^2. \tag{12}$$

Calculating the Cholesky factor, $\boldsymbol{L}(\boldsymbol{\theta})$, is the crucial step in solving the penalized least squares problem (12).

The current implementation of linear mixed models in the `lme4` package for R allows for fitting models of the form (8) to very large data sets. The rate at which state-wide data are being accumulated, however, will soon push the sizes of models past one of the inherent limitations in the R implementation related to memory size. The problem is not hardware; most new computers incorporate 64-bit processors allowing access to more than 4 gigabytes of memory and the memory modules themselves are relatively inexpensive. The problem is in adapting software to use 64-bit memory addresses. The sparse matrix code being used in the `lme4` package for R does allow for extended addresses but R itself does not yet allow this, nor is it likely to be modified to do so in the near future.

To be able to fully utilize large memory spaces to fit VAMs to very large databases may require a hybrid approach with special code to interface between the more restrictive R memory address space and the wider address space of the sparse matrix code.

3 Practical Examples of Cross-Classified Modeling for VAMs Using `lme4`

The statistical theory described previously argues for the use of a mixed-effects model in value-added modeling. But if fitting such models, even using the advanced techniques described previously, requires a substantial investment of resources, then it becomes important to evaluate if the investment is warranted. Memory is not the only limitation on fitting cross-classified VAMs to large data sets; computing time can also be an important factor, especially when performing simulations. Fitting a model to a very large data set, which could take a day or more on a powerful server computer, is tolerable when just the one fit is needed. If, however, each model fit is just one iteration in a large simulation study then more effective methods are needed.

In one of the first evaluations of different VAMs, Tekwe et al. (2004) setup their process of critical appraisal by considering four separate types of VAMs: the simple change score fixed effects model, the simple unadjusted hierarchical linear mixed model, the hierarchical linear mixed model adjusted for student- and school-level covariates, and the layered mixed effects model. The results of their study led them to prefer the simpler change score fixed effects model over more complicated models, although it should be noted that their approach did not actually provide any simulation-type estimates and was limited to observations of the models.

The models tested by Tekwe et al. (2004) have grown in their complexity, including the variable persistence model outlined by Lockwood, McCaffrey, Mariano & Setodji (2007). In addition to these five model conditions, Dallas ISD has proposed a classroom effectiveness index which has since been cited as a potential model for implementation by the Texas Education Agency (2009). The Dallas ISD model is also the only model in which data are considered in a two-stage approach where covariates are included in an ordinary least squares (OLS) setting and then the mixed-effects model is run on the residualized scores from the OLS analysis.

3.1 An Example for a Single Timepoint

```
> pupils <- read.table("pupils.txt", header = T)
> library(lme4)
```

The first model that we will fit is a model in which student achievement is modeled as a cross-classified effect of both the primary school and the secondary school that a student belongs to. Our first model will be the "null" or unconditional model such that,

$$Y_{i(j_1,j_2)} = \gamma_{000} + u_{00j_1} + u_{00j_2} + e_{i(j_1,j_2)} \tag{13}$$

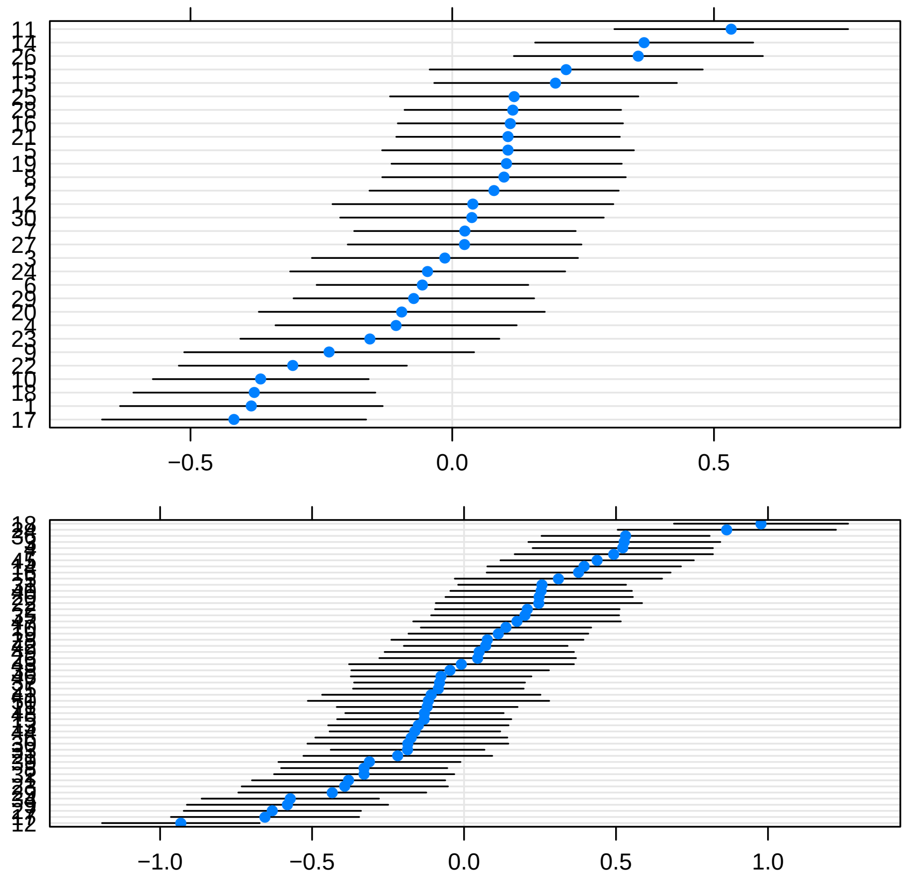
where $Y_{i(j_1,j_2)}$ is the student achievement score, u_{00j_1} is the random effect for the primary school, u_{00j_2} is the random effect for the secondary school, and $e_{i(j_1,j_2)}$ is the student-level residual.

This is easily modeled in `lme4` as:

```
> m0 <- lmer(ACHIEV ~ 1 + (1 | PSCHOOL) + (1 | SSCHOOL),
+     REML = F, pupils)
> summary(m0)

Linear mixed model fit by maximum likelihood
Formula: ACHIEV ~ 1 + (1 | PSCHOOL) + (1 | SSCHOOL)
   Data: pupils
  AIC  BIC logLik deviance REMLdev
 2326 2345  -1159     2318    2321
Random effects:
 Groups   Name        Variance Std.Dev.
 PSCHOOL  (Intercept) 0.169348 0.41152
 SSCHOOL  (Intercept) 0.065401 0.25574
 Residual             0.513169 0.71636
Number of obs: 1000, groups: PSCHOOL, 50; SSCHOOL, 30

Fixed effects:
            Estimate Std. Error t value
(Intercept)  6.34865    0.07831   81.07
```



We can see from this Figure that the variability of the Primary School scores is larger than the variability of the Secondary Schools. We can also see this in the variance components in the model summary.

In addition to model `m0`, we can test for the effect of pupil ses level and pupil gender as fixed effects such that:

$$Y_{i(j_1,j_2)} = \gamma_{000} + \gamma_{100}\text{gender}_{i(j_1,j_2)} + \gamma_{200}\text{ses}_{i(j_1,j_2)} + u_{00j_1} + u_{00j_2} + e_{i(j_1,j_2)} \qquad (14)$$

This can be modeled with:

```
> m1 <- lmer(ACHIEV ~ PUPSEX + PUPSES + (1 | PSCHOOL) +
+      (1 | SSCHOOL), REML = F, pupils)
> summary(m1)
```

```
Linear mixed model fit by maximum likelihood
Formula: ACHIEV ~ PUPSEX + PUPSES + (1 | PSCHOOL) + (1 | SSCHOOL)
   Data: pupils
  AIC  BIC logLik deviance REMLdev
 2255 2285  -1122     2243    2258
Random effects:
```

```
 Groups   Name        Variance Std.Dev.
 PSCHOOL  (Intercept) 0.169010 0.41111
 SSCHOOL  (Intercept) 0.063606 0.25220
 Residual             0.474255 0.68866
Number of obs: 1000, groups: PSCHOOL, 50; SSCHOOL, 30

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.75548    0.10527   54.67
PUPSEXgirl   0.26131    0.04564    5.73
PUPSES       0.11409    0.01610    7.09

Correlation of Fixed Effects:
           (Intr) PUPSEX
PUPSEXgirl -0.254
PUPSES     -0.643  0.075
```

Furthermore, we may wish to model random effects for `gender` across all primary schools (and not secondary schools). This would be specified as:

$$Y_{i(j_1,j_2)} = \gamma_{000} + \gamma_{100}\text{gender}_{i(j_1,j_2)} + (\gamma_{200} + u_{10j_1})\text{ses}_{i(j_1,j_2)} + u_{00j_1} + u_{00j_2} + e_{i(j_1,j_2)} \quad (15)$$

```
> m2 <- lmer(ACHIEV ~ PUPSEX + PUPSES + (PUPSEX | PSCHOOL) +
+     (1 | SSCHOOL), REML = F, pupils)
> summary(m2)

Linear mixed model fit by maximum likelihood
Formula: ACHIEV ~ PUPSEX + PUPSES + (PUPSEX | PSCHOOL) + (1 | SSCHOOL)
   Data: pupils
  AIC  BIC logLik deviance REMLdev
 2257 2296  -1120     2241    2255
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 PSCHOOL  (Intercept) 0.144265 0.37982
          PUPSEXgirl  0.016480 0.12838  0.463
 SSCHOOL  (Intercept) 0.063742 0.25247
 Residual             0.470093 0.68563
Number of obs: 1000, groups: PSCHOOL, 50; SSCHOOL, 30

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.75892    0.10278   56.03
PUPSEXgirl   0.26551    0.04900    5.42
PUPSES       0.11301    0.01608    7.03

Correlation of Fixed Effects:
```

```
           (Intr) PUPSEX
PUPSEXgirl -0.150
PUPSES     -0.658  0.069
```

3.2 An Example for Time-Series Cross Classified Models

```
> vamdata <- data.frame(time = rep(0:5, 12), id = rep(1:12,
+     each = 6), teach1 = rep(10:15, each = 12), teach2 = rep(c(20,
+     21, 20, 21, 20, 23, 23, 24, 23, 24, 21, 24), each = 6),
+     achieve = c(1, 3, 5, 6, 6, 7, 2, 5, 7, 9, 13,
+         15, 1, 1, 2, 2, 2, 3, 1, 2, 2, 3, 5, 8, 2,
+         3, 7, 7, 6, 7, 3, 7, 9, 9, 8, 7, 0, 0, 1,
+         1, 1, 0, 0, 1, 0, 1, 7, 9, 2, 4, 8, 8, 7,
+         7, 3, 5, 8, 9, 11, 11, 1, 3, 5, 6, 10, 12,
+         5, 5, 4, 6, 6, 7), vteach1 = rep(c(10, NA,
+         10, NA, 11, NA, 11, NA, 12, NA, 12, NA, 13,
+         NA, 13, NA, 14, NA, 14, NA, 15, NA, 15, NA),
+         each = 3), vteach2 = rep(c(NA, 20, NA, 21,
+         NA, 20, NA, 21, NA, 20, NA, 23, NA, 23, NA,
+         24, NA, 23, NA, 24, NA, 21, NA, 24), each = 3),
+     cteach = rep(c(10, 20, 10, 21, 11, 20, 11, 21,
+         12, 20, 12, 23, 13, 23, 13, 24, 14, 23, 14,
+         24, 15, 21, 15, 24), each = 3))
```

Suppose that we have a situation where we are doing continuous progress monitoring on students three times a year across two years. In this instance, we have 6 total measurement occasions for each student across two teacher classrooms. We would also note that this is not a true nested structure since students are randomly placed into teacher classrooms in Year 2 and this placement is not dependent on the teacher that the student had in Year 1.

```
> xtabs(~time + cteach, vamdata)

    cteach
time 10 11 12 13 14 15 20 21 23 24
   0  2  2  2  2  2  2  0  0  0  0
   1  2  2  2  2  2  2  0  0  0  0
   2  2  2  2  2  2  2  0  0  0  0
   3  0  0  0  0  0  0  3  3  3  3
   4  0  0  0  0  0  0  3  3  3  3
   5  0  0  0  0  0  0  3  3  3  3
```

Notice that each of the teachers with a first digit of “1” are only teaching students in time 0:2 and teachers with a first digit “2” are teaching students in times 3:5.

We can now specify the model with random effects at the teacher level. We specify the model this way because we assume that the effect of student achievement across time changes within teacher classrooms.

```
> vam1 <- lmer(achieve ~ time + (time | cteach) + (1 |
+     id), vamdata)
> summary(vam1)
```

```
Linear mixed model fit by REML
Formula: achieve ~ time + (time | cteach) + (1 | id)
   Data: vamdata
   AIC   BIC logLik deviance REMLdev
 289.7 305.6 -137.8    276.9   275.7
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 7.5470   2.7472
 cteach   (Intercept) 5.6589   2.3788
          time        1.1774   1.0851   -1.000
 Residual             1.0384   1.0190
Number of obs: 72, groups: id, 12; cteach, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept)   2.1763     1.1172   1.948
time          1.2613     0.3535   3.568

Correlation of Fixed Effects:
     (Intr)
time -0.674
```

Notice the correlation of -1.000 for the slope and intercept random effects within teacher. The model is degenerate because there is not enough information from 10 teachers to estimate two variances and a covariance.

We can also view the random effects for each teacher. This helps us identify teachers who have students who are performing at, above, or below the average teacher. This is, in a sense, a value-added component for a student being in a specific classroom. However, it should be noted that this is a norm-referenced value-added component.

```
> ranef(vam1)
```

```
$id
   (Intercept)
1    0.7244848
2    2.4727511
3   -3.1613238
4   -3.5311508
5    1.3405451
6    3.7347617
7   -3.4460175
8   -2.6188209
```

```
9    2.7345356
10   2.9100113
11  -0.7462058
12  -0.4135707

$cteach
   (Intercept)       time
10  -2.0531202  0.9365056
11   2.1450473 -0.9784371
12  -1.9189711  0.8753152
13   0.5772548 -0.2633077
14  -2.4467006  1.1160325
15   1.4752146 -0.6729011
20   2.0513364 -0.9356920
21  -2.9085858  1.3267158
23   3.5448827 -1.6169549
24  -0.4663579  0.2127235
```