A Model for Time-Varying Cumulative Classroom/Teacher Effects on Students’ Growth

Mark H. C. Lai
University of Cincinnati

Oi-man Kwok
Texas A&M University

Background

Multilevel growth curve analysis has been commonly used to model students’ learning trajectories while at the same time addressing the dependency in educational data. The data structure becomes more complicated with student mobility as most of the students move to a different classroom with a different teacher every year. To account for classroom/teacher effects, Raudenbush (1993) proposed a cross-classified random effects model (CCREM) where repeated measures were cross-classified by students and teachers, and teacher effects across time were treated as exchangeable. Whereas the acute-effects CCREM (Figure 1a) assumes that previous classroom membership does not affect outcome at a later time, Raudenbush and Bryk (2002) also discussed a cumulative effects CCREM (Figure 1b) that assume constant teacher effects for all subsequent measurements.

However, as raised in Cafri, Hedeker, and Aarons (2015) and McCaffrey, Lockwood, Koretz, and Hamilton (2004), such an assumption of constant teacher effects is questionable, as it is generally expected that teacher effects are weaker for a later time. A model that correctly captures the cumulative teacher effects is of interest for three reasons. First, the cumulative effect estimates themselves should be of substantive interest as they quantify the influences of teaching at a particular grade level across time. Second, as incorrect modeling of the variance-covariance structure was shown to bias random component and fixed effect standard error estimates in multilevel modeling in general (Berkhof & Kampen, 2004) and in CCREM specifically (Cafri et al., 2015; Luo & Kwok, 2009; Meyers & Beretvas, 2006), growth parameters and parameters of interest may be biased with the cumulative teacher effects assumed either zero or constant across time. Third, as cumulative...
Figure 1. Data structure with repeated measures cross-classified by students and classrooms with (a) accurate teacher/classroom effects and (b) cumulative teacher/classroom effects (in dashed lines). O = Observation. S = Student. KC = Kindergarten classroom. G1 = Grade 1. C1 = Classroom 1. G2 = Grade 2. C2 = Classroom 2.
effects imply that students’ performance and growth are products of learning from multiple teachers/classrooms, incorrect modeling of the cumulative effects may lead to biased model-based teacher effectiveness estimates and invalid uncertainty estimates (Hill, Kapitula, & Umland, 2011).

**Purpose**

In this study, we formulate a time-varying cumulative-effect CCREM (TVC-CCREM) that subsumes both the acute-effect and the cumulative-effect CCREMs. The model is illustrated and compared with other CCREM models using data from the Tennessee’s Student Teacher Achievement Ratio (STAR) project (Achilles et al., 2008). A simulation study is conducted to examine the impact of incorrectly modeling the cumulative effects as well as the sample size requirements in recovering the cumulative effect parameters.

**Model Formulation**

The TVC-CCREM can be formulated by the components at the repeated-measure (level-1), student, and classroom/teacher levels. In this study classrooms and teachers are treated as the same level and they are used exchangeably. Assuming a linear trend and that the data span across \( K \) grade levels so that each student attended \( K \) different classrooms, the \( r \)th repeated measure for student \( i \) who have attended a set of \( \{c\} = \{c_1, c_2, \ldots, c_K\} \) classrooms can be modelled as

Repeated-measure level:

\[
Y_{ti\{c\}} = \pi_{0i} + \pi_{1i} \text{TIME}_{ti} + v_{t\{c\}} + e_{ti}.
\]

Student level:

\[
\pi_{0i} = \theta_0 + u_{0i},
\]

\[
\pi_{1i} = \theta_1 + u_{1i}.
\]

Classroom level:

\[
v_{t\{c\}} = \sum_{k=1}^{K} \lambda_{kt} v_{kc}.
\]

At level-1, the outcome \( Y \) is determined by student-specific intercept (\( \pi_{0i} \)) and growth rate (\( \pi_{1i} \)) plus the time-varying classroom contribution (\( v_{t\{c\}} \)) and a within-student error term (\( e_{ti} \)); at the student-level, \( \theta_0 \) is the average outcome at \( \text{TIME} = 0 \), \( \theta_1 \) is the average growth rate, and \( u_{0i} \) and \( u_{1i} \) are the student-specific deviations in the intercept and growth rate; at the classroom level, \( v_{kc} \) represents the effect of the \( c_k \)th classroom in the \( k \)th set of classrooms, and \( \lambda_{kt} \) is the cumulative effect of the \( k \)th set of classrooms at time \( t \), with \( \lambda_{kt} = 1 \) when \( k \) denotes a current classroom and \( \lambda_{kt} = 0 \) when \( k \) denotes a future classroom at time \( t \). The variance of \( v_k \) is allowed to be different for different \( k \)s. As an example, the
outcome at grade 2 \( (t = 2) \) will have contributions from grade 1 classroom (freely estimated \( \lambda_{12} \)) and concurrent effect from grade 2 classroom \( (\lambda_{22} \) fixed to 1), but no contribution from grade 3 \( (\lambda_{32} \) fixed to 0). TVC-CCREM reduces to the acute-effect CCREM when all \( \lambda \) parameters are zero, and to the cumulative-effect CCREM when \( \lambda_{tk} = 1 \) for all cumulative and concurrent classroom effects.

**Empirical Illustration**

From the STAR project we chose the reading outcome (measured by the Stanford Achievement Test) measured from kindergartent to grade 3, and for demonstration purpose we included in our analysis only students with teacher identifiers on all four time points, resulting in a total of 11,745 observations from 3,080 students from 309, 330, 335, and 333 classrooms at kindergarten, grade 1, 2, and 3, respectively. Preliminary analyses showed that a quadratic growth model fitted the mean trajectory of reading scores better than a linear growth model, so we compared TVC-CCREM with the acute-effect CCREM with quadratic growth using Bayesian estimations with STAN (Carpenter et al., 2017). For simplicity we did not include the school effects, but they can be handled with STAN.

We used weakly informative priors (see Appendix) for the model parameters and estimated the model with four chains, each with 2,000 iterations (and half as warmup). The key parameter estimates were shown in Table 1, and the leave-one-out cross-validation information criteria (LOO-IC; Vehtari, Gelman, & Gabry, 2017) indicated TVC-CCREM \( (\text{LOO-IC} = 16,251.7) \) fitted better than the acute-effect CCREM \( (\text{LOO-IC} = 16,302.5) \). The \( \lambda \) estimates ranged between 0.226 and 0.536 and decreased for measurements farther apart, with 95% credible intervals containing neither 0 nor 1. The correlation between the teacher effect estimates under the two models was .932, but the SD for the Bayesian estimates (posterior means) based on the TVC-CCREM were much larger than those based on the cumulative-effect CCREM (.31 vs. .25; Figure 2).

**Simulation**

We conduct a simulation study on the impact of ignoring the cumulative effects and on whether the \( \lambda \) parameters can be recovered across conditions of students mobility, number of time points, students, and classrooms. Preliminary results showed that with as few as three time points, 10 students in each classroom, and 20 classrooms the cumulative effects were unbiasedly estimated with good 95% CI coverage. However, when the cumulative effects were ignored, the posterior SD of the fixed effects were generally underestimated, with empirical coverage rates slightly below nominal level. Full simulation results will be included in the final presentation.

**Conclusions**

Given the interest in teacher effectiveness not just on short-term outcome but also long-term growth, models that quantify cumulative teacher effects are much needed. With our proposed model we showed with real data that the cumulative effects were non-zero
Table 1
Model Results of the Empirical Illustration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TVC-CCREM</th>
<th>Acute-Effect CCREM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SD</td>
</tr>
<tr>
<td>$\theta_0$</td>
<td>8.88</td>
<td>0.0215</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>2.13</td>
<td>0.032</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>−0.313</td>
<td>0.0096</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>0.39</td>
<td>0.00396</td>
</tr>
<tr>
<td>$\sigma_{u_0}$</td>
<td>0.411</td>
<td>0.0104</td>
</tr>
<tr>
<td>$\sigma_{u_1}$</td>
<td>0.458</td>
<td>0.015</td>
</tr>
<tr>
<td>$\sigma_{u_2}$</td>
<td>0.134</td>
<td>0.00458</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>0.32</td>
<td>0.0175</td>
</tr>
<tr>
<td>$\sigma_{v_2}$</td>
<td>0.528</td>
<td>0.025</td>
</tr>
<tr>
<td>$\sigma_{v_3}$</td>
<td>0.29</td>
<td>0.0167</td>
</tr>
<tr>
<td>$\sigma_{v_4}$</td>
<td>0.206</td>
<td>0.0135</td>
</tr>
<tr>
<td>$\lambda_{12}$</td>
<td>0.536</td>
<td>0.0807</td>
</tr>
<tr>
<td>$\lambda_{13}$</td>
<td>0.295</td>
<td>0.0703</td>
</tr>
<tr>
<td>$\lambda_{14}$</td>
<td>0.226</td>
<td>0.0607</td>
</tr>
<tr>
<td>$\lambda_{23}$</td>
<td>0.477</td>
<td>0.0358</td>
</tr>
<tr>
<td>$\lambda_{24}$</td>
<td>0.296</td>
<td>0.0318</td>
</tr>
<tr>
<td>$\lambda_{34}$</td>
<td>0.392</td>
<td>0.0568</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>0.299</td>
<td>0.00891</td>
</tr>
</tbody>
</table>

Note. The outcome variable is reading score (divided by 50 for better computational stability). $\theta_1$ and $\theta_2$ are the fixed effects for linear and quadratic growth. TVC = time-varying cumulative effects. CCREM = cross-classified random effects model. Est = posterior median as the Bayesian point estimate. CI = equal-tailed credible interval.

even after three years, but also were much smaller than one as assumed in the constant-cumulative-effect CCREM. The model can be easily expanded by incorporating intervention variables and covariates at different levels, and intervention effects and group differences on the cumulative effects may be of particular interest.
Figure 2. Posterior means of classroom effects in the empirical illustration.

References


Appendix

Priors Used in the Empirical Illustration

Independent priors for the following parameters were specified:

\[ \theta_0 \sim U(-\infty, \infty), \]
\[ \theta_p \sim N(0, 10), p = 1, 2, \]
\[ \sigma_e \sim t_3^+(0, 10), \]
\[ \sigma_{u_p} \sim t_3^+(0, 10), p = 0, 1, 2, \]
\[ \sigma_{v_t} \sim t_3^+(0, 10), t = 1, 2, 3, 4, \]
\[ \lambda_{kt} \sim N(0, 2), k = 1, 2, 3; t = k + 1, \ldots, 4, \]

where \( N(\mu, \sigma) \) denotes a normal distribution with mean \( \mu \) and \textit{standard deviation} \( \sigma \) (to be consistent with STAN), and \( t_\nu^+(\mu, \sigma) \) denotes a non-standardized half-folded Student’s \( t \) distribution with location parameter \( \mu \), scale parameter \( \sigma \), and degrees of freedom \( \nu \).