Incorporating Mobility in Growth Modeling for Multilevel and Longitudinal Item Response Data

In-Hee Choi and Mark Wilson

aKorean Educational Development Institute; bUniversity of California, Berkeley

ABSTRACT

Multilevel data often cannot be represented by the strict form of hierarchy typically assumed in multilevel modeling. A common example is the case in which subjects change their group membership in longitudinal studies (e.g., students transfer schools; employees transition between different departments). In this study, cross-classified and multiple membership models for multilevel and longitudinal item response data (CCMM-MLIRD) are developed to incorporate such mobility, focusing on students’ school change in large-scale longitudinal studies. Furthermore, we investigate the effect of incorrectly modeling school membership in the analysis of multilevel and longitudinal item response data. Two types of school mobility are described, and corresponding models are specified. Results of the simulation studies suggested that appropriate modeling of the two types of school mobility using the CCMM-MLIRD yielded good recovery of the parameters and improvement over models that did not incorporate mobility properly. In addition, the consequences of incorrectly modeling the school effects on the variance estimates of the random effects and the standard errors of the fixed effects depended upon mobility patterns and model specifications. Two sets of large-scale longitudinal data are analyzed to illustrate applications of the CCMM-MLIRD for each type of school mobility.

Introduction

Multilevel models or hierarchical linear models (e.g., Goldstein, 2003; Raudenbush & Bryk, 2002) are commonly used to handle nested data structures in behavioral science and other application areas. One of the standard assumptions in multilevel modeling is that the data structure is strictly hierarchical; for example, schools are nested within neighborhoods. However, multilevel data often cannot be represented by the strict form of hierarchy. A common example is the case in which subjects change their group membership in longitudinal studies (e.g., employees transition to different departments within a company; patients transfer to different hospitals; children transfer schools).

In particular, in educational research, large-scale longitudinal studies (e.g., High School and Beyond [HS&B], National Education Longitudinal Study of 1988 [NELS:88] and Early Childhood Longitudinal Studies [ECLS]) have provided plenty of resources for researchers not only to examine individual growth of students over time but also to investigate the effect of contextual factors, such as teacher education and types of school, on student growth. A major circumstance that causes complex data structures in large-scale longitudinal data is group membership change: Some students move from one school to another over the course of the data collection, referred to as school mobility in the present study. In this article, two types of school mobility frequently observed in large-scale longitudinal assessments (Luo & Kwok, 2012) are described, and appropriate approaches to analyze the resulting data are developed.

The first type of school mobility corresponds to the simultaneous movement of students at a certain time point due to promotion within the education system; for example, middle school graduation and high school matriculation (e.g., in the NELS:88 and the Korean Youth Panel Survey [KYPs]). In this type of mobility, each student has a combined membership of middle school and high school. The data structure in this example is represented in Figure 1, part (a), in which rectangles represent sets of classification units, and arrows going from the lower-level to the higher-level units describe membership classifications. In cross-classified models developed to analyze this type of multilevel data (Rasbash & Goldstein, 1994; Raudenbush, 1993), two classifications at Level 2 for the middle and high school (e.g., two separate rectangles in part [a] of Figure 1) are assumed, and the students have a membership in each classification (e.g., one arrow from the stu-
In part (b) of Figure 1, the school is a single classification unit at Level 2, represented by a rectangle, and the double arrows from the student to the school display the student’s multiple school membership within the school level. Source: Browne et al. (2001, p. 106). Adopted with permission.

Another pattern of school mobility occurs when subsamples of students switch school or classroom membership. For example, students can transfer to other schools for various reasons such as moving, parents’ job change, or other issues during the repeated measurement occasions. In such cases, some of the students can move at any time during the data collection, and it is possible that they switch school membership multiple times. A report by the U.S. Government Accounting Office (U.S. Government Accounting Office, 1994) showed that the average mobility rate, defined as the percentage of students who switched schools, was 17%, and for some populations, the rates were much higher; for example, as high as 40% (Grady & Beretvas, 2010). This type of data structure is addressed in multiple membership models (Hill & Goldstein, 1998; Rasbash & Browne, 2001), in which lower-level units are simultaneously members of more than one unit within the same higher-level classification. In part (b) of Figure 1, the school is a single classification unit at Level 2, represented by a rectangle, and the double arrows from the student to the school display the student’s multiple school membership within the school level.

The outcomes of interest in the present study are multilevel and longitudinal test data from large-scale assessments, in which responses on the same set of items from the same students are collected over time. Thus, responses on items (Level 1) are clustered into a timepoint (Level 2), and repeated occasions are nested within a student (Level 3). When the two types of school mobility are observed in multilevel and longitudinal item response data, the cross-classified and multiple-membership models can be used to extend the three-level model in order to incorporate school effects in growth modeling. In the proposed models, the item responses can be directly related to a latent variable via an IRT model. For illustrative purposes, in this study, only the Rasch model and binary outcomes are considered. A benefit of using the IRT model with repeated item response data is that item and person parameters are estimated simultaneously (referred to as a one-step analysis). When the IRT scaled scores are used as the dependent variables as in previous studies (e.g., Grady & Beretvas, 2010; Luo & Kwok, 2012; Palardy, 2010), a two-step procedure is taken: Person measures are obtained first by fitting an IRT model to the data (Stage 1), and then estimated person measures are applied to growth modeling (Stage 2). A limitation of this approach is that person measures are treated as true scores without measurement errors, which can lead to incorrect results in the subsequent analysis, especially when the number of items in the test is not large and, hence, measurement error is not negligible (Adams, Wilson, & Wu, 1997; Hung & Wang, 2012; Kamata, 2001). Moreover, compared to the total approach in which an unweighted mean computed from responses on multiple measures is generally used, incomplete designs and missing responses are no longer obstacles to obtain person measures in the IRT models.
In summary, the proposed models will take advantage of item response models and cross-classified and multiple membership models. In a recent study by Kelcey, McGinn, and Hill (2014), cross-classified IRT models were proposed to model cross-classified dependence in repeated rater-mediated assessments. When each observation is rated by different raters across timepoints, each observation is cross-classified by a participant and a rater. Along with this study, the current article can be considered as extensions of cross-classified and multiple membership models into IRT, but our focus is on modeling effects of multiple schools in growth modeling, which reveals how individual students change over time and how schools influence student growth.

Another goal of the current study is to investigate the effect of incorrectly modeling school membership in the analysis of multilevel and longitudinal item response data. When mobile students are found in longitudinal data, one option for researchers who rely solely on the traditional multilevel models is to ignore school membership and use the three-level model. In this case, unobserved school effects, shared by the students who attended the same schools, are not modeled properly. Another possible option is to assume that the students stay within the same schools; for example, by using only the information for the first or last school they attended. In real settings, because large-scale longitudinal studies are observational studies in most cases, the school information of students who switch schools may not be fully reported. Therefore, the possible effects of multiple schools on students who have attended more than one school are excluded in this approach. Again, the consequences of such an incorrect specification have been investigated in the cross-classified and multiple membership literature for the analysis of cross-sectional data (e.g., Chung & Beretvas, 2011; Luo & Kwok, 2009; Meyers & Beretvas, 2006) and longitudinal continuous outcomes (e.g., Grady & Beretvas, 2010; Luo & Kwok, 2012). The findings from previous studies showed that incorrectly modeling school membership yielded biased variance estimates for the random effects and resulted in biased standard errors for the fixed effects. The current study investigates whether the previous findings are applicable to the analysis of multilevel and longitudinal item response data.

This article is organized as follows. First, traditional approaches to longitudinal item response data are introduced. Second, models are developed to deal with the two types of school mobility according to the cross-classified and multiple membership models, and a brief explanation is given as to how the Bayesian methods can be employed to fit the proposed models. Third, to assess parameter recovery and the effect of incorrect school membership, two simulation studies for the two types of school mobility are conducted. Note that the simulation studies are designed to mimic the scenario of large-scale longitudinal assessments that involve large sample sizes. Finally, empirical examples of real data sets from large-scale longitudinal studies are illustrated.

**Methods**

**Hierarchical generalized linear model for multilevel and longitudinal item response data (HGLM-MLIRD)**

The measurement of individual growth or change in a construct is a focus of studies in educational and psychological research settings. For investigating growth, the same set of items (or with common items at least) is administered to the students repeatedly over time, and longitudinal item response data are collected. While students are measured repeatedly with items in typical two-level item response models (e.g., Adams et al., 1997; Mislevy & Bock, 1989), in longitudinal item response data, students are measured repeatedly in two aspects: measurement occasions and items (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006). This allows the employment of three-level modeling. In detail, a set of responses from a student on one occasion are more alike than responses from another occasion, and responses from the same student are more correlated than those from another student. As a consequence, there are two possible types of within-cluster correlations in longitudinal item response data: (1) within-student and within-occasion correlation and (2) within-student and between-occasion correlation. To deal with different within-student correlations, a three-level approach is needed, in which item responses are nested within an occasion and occasions are nested within students (Pastor & Beretvas, 2006; Segawa, 2005). The three-level approach is specified in the following models.

**The Level 1 model**

The Level 1 model, referred to as the measurement model, specifies the item response functions. Let $i_{yt}$ denote the response to item $i$ at measurement occasion $t$ for student $j$, for $i = 1, \ldots, I$, $t = 1, \ldots, T$, and $j = 1, \ldots, J$. Using the Rasch model, the probability that student $j$ gives a correct response on item $i$ at occasion $t$ is written as follows:

$$
\log \left( \frac{P(y_{it} = 1 | \theta_{ij})}{1 - P(y_{it} = 1 | \theta_{ij})} \right) = \theta_{ij} + \sum_{q=1}^{1} X_{qit} \delta_{iq},
$$

where $\delta_{iq}$ represents the latent variable of student $j$ at occasion $t$, $X_{qit}$ is the $q$th indicator variable with value of $-1$.
when \( q = i \) and 0 when \( q \neq i \); and \( \delta_i \) denotes the difficulty parameter of item \( i \). The latent variable \( \theta_{tj} \) is occasion specific as well as student specific, indicating it is a time-varying variable. In Equation (1), item difficulties are fixed to be invariant across measurement occasions, with the constraint \( \delta_i = -\sum_{i=1}^{T-1} \delta_i \).

**The Level 2 model**

At Level 2, a latent growth curve model (Duncan, Duncan, & Strycker, 2006; McArdle & Epstein, 1987) is specified to demonstrate the latent variable of student \( j \) at occasion \( t \) as a function of the time variable, allowing for estimation of individual growth trajectories. For example, the Level 2 (between-occasion and within-student) model for the latent variable of the Level 1 model can be expressed as a linear growth model,

\[
\theta_{tj} = \pi_0j + \pi_1j d_i + \varepsilon_{tj}, \tag{2}
\]

where \( d_i \) is the time variable taking on values of 0, 1, \ldots, \( T-1 \) for occasion 1, 2, \ldots, \( T \). In Equation (2), \( \pi_0j + \pi_1j d_i \) is the linear growth trajectory of student \( j \), where \( \pi_0j \) and \( \pi_1j \) represent the initial status (intercept) and linear change (slope) of the latent variable, and \( \varepsilon_{tj} \) is the deviation (residual) at occasion \( t \) from the linear growth trajectory of student \( j \). In the growth model, \( \varepsilon_{tj} \) is often assumed to be normally distributed with mean zero and a constant variance, \( \varepsilon_{tj} \sim N(0, \sigma^2) \); that is, an independent and identically distributed (i.i.d.) structure. As an extension of the linear growth model, higher order polynomials of the time variable and time-varying covariates can be included, and it is possible to assume an alternative specification of \( \varepsilon_{tj} \), such as an autoregressive structure (Hung & Wang, 2012; Segawa, 2005).

**The Level 3 model**

In the Level 3 (between-student) model, the student-specific growth parameters serve as dependent variables,

\[
\pi_{0j} = \beta_0 + \zeta_{0j}, \quad \pi_{1j} = \beta_1, \tag{3}
\]

where \( \beta_0 \) and \( \beta_1 \) are the fixed intercept and linear growth rate across students, respectively, and \( \zeta_{0j} \) is the random effect (intercept or residual) of student \( j \). It is assumed that \( \zeta_{0j} \) follows a normal distribution, \( \zeta_{0j} \sim N(0, \psi^2) \), and \( \text{Cov}(\zeta_{0j}, \varepsilon_{tj}) = 0 \). Substituting Equation (3) into Equation (2) yields the latent regression (Adams et al., 1997) for \( \theta_{tj} \),

\[
\theta_{tj} = \beta_0 + \beta_1 d_i + \zeta_{0j} + \varepsilon_{tj}. \tag{4}
\]

In addition, Equation (4) can be rewritten in a matrix for student \( j \) as follows:

\[
\Theta_j = X_j B + Z_j \zeta_{0j} + \varepsilon_j, \tag{5}
\]

where \( \Theta_j = \begin{bmatrix} \theta_{tj} \\ \vdots \\ \theta_{Tj} \end{bmatrix} \), \( X_j = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & T-1 \end{bmatrix} \), \( B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \), \( Z_j = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \), and \( \varepsilon_j = \begin{bmatrix} \varepsilon_{tj} \\ \vdots \\ \varepsilon_{Tj} \end{bmatrix} \).

**Two types of school mobility in multilevel and longitudinal item response data**

As discussed earlier, when students switch schools in the course of repeated measurements in longitudinal studies, conventional multilevel modeling is not applicable. In this article, cross-classified and multiple membership models for multilevel and longitudinal item response data (CMM-MLIRD) are developed to deal with the two types of school mobility that are often observed in large-scale longitudinal assessments.

**Type 1**

The first type of school mobility describes the simultaneous movement of students at a certain timepoint due to promotion within the education system. For example, in the Korean Youth Panel Survey (KYPS; National Youth Policy Institute, 2009), the first survey was administered to second-year middle school students, and researchers followed them once a year until high school graduation. Across five occasions, students moved from middle schools to high schools between the second and third occasions. In this case, the strict three-level data structure must be extended to the cross-classified models, in which students are nested within a middle and high school combination.

To illustrate, in Figure 2, which is similar to one suggested by Jeon and Rabe-Hesketh (2012), solid rectangles and arrows represent a clustered structure: items, occasions, and students. In particular, middle and high schools are represented as separate and unconnected rectangles located at the same level, and the cross-classified relationship is described by two arrows extending from a student to a middle or high school. Furthermore, dotted rectangles indicate specific timepoints within the time level. Suppose that a student attended a middle school at timepoint \( t' \) and a high school at timepoint \( t'' \). Therefore, the responses at time \( t' \) are nested into the middle school, and those at time \( t'' \) are nested into the high school, represented by dotted arrows.

The distinguishing characteristic of this type of mobility is that students switch schools at the same time, separating measurement occasions into the two distinct periods (e.g., years of middle and high school). To investigate different growth patterns during middle and high
school, a piecewise growth model that allows for splitting the growth trajectories into several linear components according to distinct developmental periods is used (Li, Duncan, Duncan, & Hops, 2001; Raudenbush & Bryk, 2002). For example, in the KYPS example, for the two-piece linear growth model, two time-related variables, \( d_{1t} \) and \( d_{2t} \), are composed using the coding scheme in Table 1, and the coefficients of \( d_{1t} \) and \( d_{2t} \) present the growth rate (slope) while attending middle school and high school, respectively.

Suppose that there are \( M \) middle schools and \( H \) high schools, and the middle and high schools are indexed by \( m = 1, \ldots, M \) and \( h = 1 \ldots, H \). The response on item \( i \) at occasion \( t \) of student \( j \) who attended middle school \( m \) and high school \( h \) is denoted by \( y_{ijtmh} \), and the Level 1 measurement model is written as

\[
\log\left( \frac{P(y_{ijtmh} = 1|\theta_{ijtmh})}{1 - P(y_{ijtmh} = 1|\theta_{ijtmh})} \right) = \theta_{ijtmh} + \sum_{q=1}^{l} X_{qij} \delta_{q},
\]

where \( \theta_{ijtmh} \) is the latent variable at occasion \( t \) of student \( j \) who attended middle school \( m \) and high school \( h \); \( X_q \) is the \( q \)th indicator variable with value of \(-1\) when \( q = i \) and \( 0 \) when \( q \neq i \); and \( \delta_q \) indicates the fixed difficulty parameter of item \( q \). In the adoption of the two-piece linear growth model, \( \theta_{ijtmh} \) is written in the reduced form of the latent variable as follows:

\[
\theta_{ijtmh} = \beta_0 + \beta_1 d_{1t} + \beta_2 d_{2t} + \xi_{0j} + w_{1t} \gamma_{0m} + w_{2t} \eta_{0h} + \epsilon_{ijtmh},
\]

where \( \beta_0 \) is the fixed intercept; \( \beta_1 \) and \( \beta_2 \) represent the fixed slopes while attending middle school and high school, respectively; \( \xi_{0j} \) denotes the random effect of student \( j \) related to the intercept; and \( \epsilon_{ijtmh} \) is the residual at Level 2. To explain the deviations from a student-specific growth line due to student \( j \)'s studying in middle school \( m \) and high school \( h \), the school-specific random effects, \( \gamma_{0m} \) and \( \eta_{0h} \), related to the intercepts for middle and high schools, respectively, are specified. In addition, \( w_{1t} \) and \( w_{2t} \) are the coefficients that associate the middle and high school effects with the latent variable at a specific time-point \( t \).

Equation (7) can be rewritten in a matrix form as follows:

\[
\Theta_{jtmh} = X_t B + Z_t \xi_{0j} + Z_2 \gamma_{0m} + Z_3 \eta_{0h} + \epsilon_{jtmh},
\]

where \( \Theta_{jtmh} = \begin{bmatrix} \theta_{jtmh} \\ \epsilon_{jtmh} \end{bmatrix} \), \( X_t = \begin{bmatrix} 1 & d_{1t} & d_{2t} \\ \vdots & \vdots & \vdots \end{bmatrix} \), \( B = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 \end{bmatrix} \), \( Z_1 = \begin{bmatrix} \xi_{0j} \\ \vdots \end{bmatrix} \), \( Z_2 = \begin{bmatrix} w_{1t} \\ \vdots \end{bmatrix} \), \( Z_3 = \begin{bmatrix} \gamma_{0m} \\ \eta_{0h} \end{bmatrix} \), and \( \epsilon_{jtmh} = \begin{bmatrix} \epsilon_{jtmh} \\ \vdots \end{bmatrix} \).

The student-level and school-level random intercepts are assumed to follow a normal distribution with mean zero and a constant variance; \( \xi_{0j} \sim N(0, \psi_2^2) \), \( \gamma_{0m} \sim N(0, \tau_1^2) \), and \( \eta_{0h} \sim N(0, \tau_2^2) \). Thus, \( \tau_1^2 \) and \( \tau_2^2 \) indicate the variances of the random effects of middle and high schools, respectively. As in the three-level model, the Level 2 residual is assumed to follow a normal distribution with a constant variance: \( \epsilon_{jtmh} \sim N(0, \sigma^2) \). It is further assumed that \( \text{Cov}(\xi_{0j}, \gamma_{0m}) = \text{Cov}(\xi_{0j}, \eta_{0h}) = \text{Cov}(\epsilon_{jtmh}, \gamma_{0m}) = \text{Cov}(\eta_{0h}, \epsilon_{jtmh}) = 0 \).

---

**Table 1.** An example of coding scheme for the two-piece linear growth model.

|       | \( t = 1 \) | \( t = 2 \) | \( t = 3 \) | \( t = 4 \) | \( t = 5 \) |
|-------|-------------|-------------|-------------|-------------|-------------|
| \( d_{1t} \) | 0           | 1           | 2           | 2           | 2           |
| \( d_{2t} \) | 0           | 0           | 0           | 1           | 2           |

Note. \( t \) represents time (e.g., \( t = 1 \) for the initial timepoint); \( d_{1t} \) and \( d_{2t} \) are time-related variables that split the growth trajectories into two linear components (i.e., \( d_{1t} \) for middle school and \( d_{2t} \) for high school).
The coefficients \(w_{11}\) and \(w_{23}\) can be preassigned values or unknown parameters that are freely estimated (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). In the KYPS, if the school effects were constant over time and assuming that the middle schools did not affect students’ responses when they were in the high schools, then \(Z_2 = (1, 1, 0, 0, 0)'\) and \(Z_3 = (0, 0, 1, 1, 1)'\). In addition, the cumulative effects of middle school can be specified using vectors, \(Z_2 = (1, 1, 1, 1, 1)'\) and \(Z_3 = (0, 0, 1, 1, 1)'\). In addition, the assumption of the constant school effects can be relaxed by allowing estimation of the varied effects of schools using vectors \(Z_2 = (1, w_{12}, w_{13}, w_{14}, w_{15})'\) and \(Z_3 = (0, 0, 1, w_{24}, w_{25})'\). The coefficients \(w_{11}\) and \(w_{23}\) are set to a value of one for model identification, and \(w_{21}\) and \(w_{22}\) are fixed to zero because the students were in the middle school at those timepoints. In this case, \(w_{11}\) and \(w_{23}\) are the coefficients of unknown constants associated with the random effects. They represent the relationship of the responses to the underlying latent variables (i.e., factors)—more specifically, how the random effects of the middle and high schools influence observed responses on items—and are thus referred to as factor loadings (for more details, see Jeon & Rabe-Hesketh, 2012; McArdle, 1998; Rabe-Hesketh, Skrondal, & Pickles, 2004). Because \(w_{11}\) and \(w_{23}\) are both set equal to one, when the coefficients (i.e., \(w_{12}, w_{13}, w_{14}, w_{15}, w_{24}, \) and \(w_{25}\)) are estimated to be greater than that one, this indicates that the school effects increase compared to the initial timepoint (e.g., \(t = 1\) for middle schools, and \(t = 3\) for high schools).

**Type II**

The second type of school mobility corresponds to multiple school membership. Consider an example of student achievement measured annually for three years, in which the mobility rate is about 20%. One group (comprising most of the students) remains within the same school over time. Another group of students (a mobile group) consists of those who switched schools once between occasions 1 and 2 or between occasions 2 and 3. A third group consists of students who changed schools at both occasions 2 and 3. As a consequence, in this scenario, the students who attend more than one school have been under the influence of multiple schools. Unlike Type I mobility, the schools are located in one cluster, represented by a solid rectangle, and the particular schools within the school level are displayed by small dotted rectangles (see Figure 3). The students’ membership of multiple schools is expressed using double solid arrows from the student to the school as in part (b) of Figure 1. In addition, dotted arrows show the nested relationship such as item responses at occasion \(t'\) of student \(j\) into school \(s'\).

In the present study, to model students’ multiple school membership, a notation suggested by Browne, Goldstein, and Rasbash (2001) is used, and the schools that student \(j\) has attended across occasions are denoted by \(s(j)\). Let \(S\) denote the total number of schools with \(s(j)\) as a subset of the full set of schools: \(s(j) \in \{1, \ldots, S\}\). For example, in the case of Figure 3, student \(j\) attended two schools, \(s'\) and \(s''\), so \(s(j) = \{s', s''\}\). Then, the response to item \(i\) at measurement occasion \(t\) of student \(j\) who has attended schools \(s(j)\) is written as \(y_{itj(s)}\). The probability of a correct response is specified as follows:

\[
\log \left( \frac{P(y_{itj(s)} = 1 | \theta_{itj(s)})}{1 - P(y_{itj(s)} = 1 | \theta_{itj(s)})} \right) = \theta_{itj(s)} + \sum_{q=1}^{t} X_{qi} \delta_{q},
\]

where \(\theta_{itj(s)}\) is the latent variable at occasion \(t\) of student \(j\) who has attended schools \(s(j)\); \(X_{qi}\) is the \(q\)th indicator variable with value of \(-1\) when \(q = i\) and \(0\) when \(q \neq i\).
where $\beta_0$ and $\beta_1$ are the fixed intercept and linear slope of the linear growth line, respectively; $\zeta_{0j}$ is the random intercept of student $j$; $\lambda_{tjk}$ is the preassigned coefficient for student $j$ who attended school $k$ at time $t$; $v_{0k}$ is the random effect of school $k$; and $e_{tjs(j)}$ is the Level 2 residual. The random intercepts of the student and school are assumed to follow a normal distribution with mean zero and a constant variance, $\zeta_{0j} \sim N(0, \psi^2)$ and $v_{0k} \sim N(0, \tau^2)$, where $\psi^2$ and $\tau^2$ represent the between-student and between-school variance, respectively. A constant variance is specified for $e_{tjs(j)}$, which follows a normal distribution, $e_{tjs(j)} \sim N(0, \sigma^2)$. The random effects of the student and the school are independent of each other, and the Level 2 residual is independent of the random effects of the student and the school; that is, $\text{Cov}(\zeta_{0j}, v_{0k}) = \text{Cov}(\zeta_{0j}, e_{tjs(j)}) = \text{Cov}(v_{0k}, e_{tjs(j)}) = 0$.

Equation (10) can be rewritten in a matrix for student $j$ as follows:

$$\Theta_{ji(j)} = X_i B + Z_{i1} \zeta_{0j} + Z_{i2} v_{0i(j)} + e_{tjs(j)}, \quad (11)$$

where $$\Theta_{ji(j)} = \begin{bmatrix} \theta_{tjs(j)} \\ \theta_{tjs(j)} \end{bmatrix}, \quad X_i = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ \vdots & 1 & \ldots & \vdots \\ 1 & 0 & \ldots & 1 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix},$$

$$Z_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

and $$e_{tjs(j)} = \begin{bmatrix} e_{tjs(j)} \\ e_{tjs(j)} \end{bmatrix}.$$

Furthermore, $v_{0i(j)}$ is the vector of the random effects of schools student $j$ attended, and $Z_{i2}$ is the matrix of the coefficients $\lambda_{tjk}$; hence, their specifications depend on students’ school mobility patterns. For example, for student $j$ who attended two schools over three timepoints, such as school 1 at occasions 1 and 2 and school 2 at occasion 3, $v_{0i(j)} = \begin{bmatrix} v_{01} \\ v_{02} \end{bmatrix}$ and $Z_{i2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$. The coefficient $\lambda_{tjk}$ indicates the proportion of time that student $j$ attended school $k$ up to occasion $t$; thus, $\Sigma_{k \in s(j)} \lambda_{tjk} = 1$ for each timepoint $t$ (for each row of $Z_{i2}$). Likewise, if he or she switched schools two times, such as school 1 at occasion 1, school 2 at occasion 2, and school 3 at occasion 3, $v_{0i(j)} = \begin{bmatrix} v_{01} \\ v_{02} \\ v_{03} \end{bmatrix}$ and $Z_{i2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$. For students who remained in the same school (e.g., school 1), $v_{0i(j)} = v_{01}$ and $Z_{i2} = 1$.

**Estimation**

Bayesian estimation using Markov Chain Monte Carlo (MCMC) was implemented in the WinBUGS 1.4.3. software (Lunn, Thomas, Best, & Spiegelhalter, 2000) to estimate the parameters of the proposed models in the current study. Bayesian methods have been widely adopted for fitting cross-classified effects and multiple membership models (Chung & Beretvas, 2011; Grady & Beretvas, 2010; Lockwood, McCaffrey, Mariano, & Setodji, 2007). In particular, Bayesian estimation using MCMC enables extensions of cross-classified and multiple membership models into more general settings including categorical dependent variables and structural equation modeling (Asparouhov & Muthén, 2012; Browne et al., 2001). In the current study, due to the complexity of the model structures (e.g., discrete responses, longitudinal data, and complex nested structures of the students and the schools), MCMC was chosen for the estimation method. Furthermore, the flexibility of the WinBUGS program allows for the incorporation of various design matrices associated with the fixed and random effects, $X$ and $Z$, in the proposed model formulations.

For implementation in WinBUGS, prior distributions must be specified for all parameters, which include item difficulties ($\delta$), growth parameters ($\beta$), coefficients of the school effects ($w$), the time-level residual variance ($\sigma^2$), the student-level residual variance ($\psi^2$), and the school-level residual variances ($\tau_1^2$ and $\tau_2^2$). Although a number of different prior distributions can be chosen, this study limits its scope to the simple and commonly used ones, such as the conjugate priors that allow the posterior distribution to belong to the same family as the prior distributions. More specifically, assuming a normal distribution is standard practice for the growth and item parameters, and the conjugate prior for the variance of the normal distribution is the inverse-gamma distribution (Cho & Cohen, 2013; Cohen & Bolt, 2005; Gelman et al., 2014). Specifically, the prior and hyperprior distributions for the CCMM-MLIRD for Type I mobility were specified as follows:

$$\delta_i \sim N(0, 1), i = 0, \ldots, I - 1,$$
$$\beta_k \sim N(0, 1), k = 0, 1, 2,$$
$$w_{1t} \sim N(0, 10^3), t = 2, 3, 4, 5,$$
$$w_{2t} \sim N(0, 10^3), t = 4, 5,$$
$$\varepsilon_{t, j, Nh} | \sigma^2 \sim \text{Inverse} - \text{Gamma}(10^{-3}, 10^{-3}),$$
$$\sigma^2 \sim \text{Inverse} - \text{Gamma}(10^{-3}, 10^{-3}),$$
$$\zeta_{0j} | \psi^2 \sim N(0, \psi^2),$$
$$\psi^2 \sim \text{Inverse} - \text{Gamma}(10^{-3}, 10^{-3}),$$
$$\gamma_{on} | \tau_1 \sim N(0, \tau_1^2).$$
Note that diffuse priors were specified for the coefficients of the school effects and the variances of the random effects. For the item difficulties ($\delta$) and growth parameters ($\beta$), a mildly informative prior, a normal distribution with mean 0 and variance 1, was set to make the fitting procedures more stable by providing rough bounds on the model parameters (Bolt, Cohen, & Wollack, 2002; Cho & Cohen, 2010).

For all of the models considered in this study, three chains with dispersed starting values were run with 5,000 post-burn-in iterations after 5,000 iterations of burn-in. Convergence of the three chains was examined using the posterior predictive $\hat{R}$ index proposed by Gelman and Rubin (1992) with a critical value of 1.01. The deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002), which is a fit index in Bayesian estimation, was used to compare model fit. In addition, for the absolute model data fit in empirical data analyses, posterior predictive model checking (PPMC; Gelman et al., 2014), which is a Bayesian data-model-fit-checking technique to assess the plausibility of posterior predictive replicated data against observed data, was evaluated. Let $y$ be the observed data and $y^{rep}$ be the replicated data. A test statistic $T$ is chosen to detect the systematic discrepancy between $y$ and $y^{rep}$. The posterior predictive $p$ value is the comparison between two test statistics,

$$p = \Pr[T(y^{rep}) \geq T(y)|y].$$  \hspace{1cm} (13)

An extreme $p$ value (close to 0 or 1) indicates a poor model data fit. Sinharay, Johnson, and Stern (2006) found that the item pair odds ratios (OR) was an adequate test statistic for detecting misfit in standard IRT models. The OR is given by

$$OR = \frac{n_{i1}n_{00}}{n_{10}n_{01}},$$  \hspace{1cm} (14)

where $n_{ikj}$ is the number of persons whose score on item $i$ is $k$ and score on item $j$ is $k'$ ($k, k' = 0$ (incorrect), 1 (correct), and $i \neq j$ for each pair of item $i$ and item $j$). The model with the smallest proportion of extreme PPMC $p$ values (i.e., less than .05 and greater than .95) represents a better fit to the data.

### Simulation study

#### Type I mobility

**Data generation**

To simulate data with Type I mobility, students were assumed to have moved from middle to high school between occasions 2 and 3 over five occasions as in the KYPS. The data were generated using the CCMM-MLIRD for Type I mobility with the two-piece linear growth model, Equations (6) and (7). The number of items ($I$) and measurement occasions ($T$) were set as 10 and 5, respectively, and the two time-related variables, $d_{i1}$ and $d_{z2}$, took on the values as described previously. The Level 2 residual $\varepsilon_{itimh}$ was generated from a normal distribution with mean 0 and variance 0.4 ($\sigma^2 = 0.4$). The student-specific random effect, $\xi_{0j}$, was generated from a normal distribution with mean 0 and variance 0.2 ($\eta_j = 0.2$). The random effect of middle school, $\gamma_{0m}$, was generated to be normally distributed with mean 0 and variance 0.2 ($\tau_i^2 = 0.2$). Likewise, the random effect of high school, $\eta_{0h}$, was generated from a normal distribution with mean 0 and variance 0.2 ($\tau_i^2 = 0.2$), independent of $\xi_{0j}$ and $\gamma_{0m}$. These values of the variance of the random effects were adopted from previous simulation studies for the cross-classified and multiple membership models (e.g., Luo & Kwok, 2012; Meyers & Beretvas, 2006). In particular, the variance of 0.2 corresponds to a “medium” size of the variance of the student- and school-level random effects according to the criteria used by Meyers and Beretvas (2006) and Raudenbush and Liu (2001). For example, the intraclass correlation (ICC) for middle schools, computed by $\tau_i^2/ (\sigma^2 + \psi^2 + \tau_i^2 + \tau_j^2)$, is 0.2. In addition, the following coefficients for the school effects were specified: decreasing effects of middle school, $Z_2 = (1, 0.8, 0.6, 0.4, 0.2')$, and increasing effects of high school, $Z_3 = (0, 0, 1, 1.2, 1.4')$. The fixed intercept and slopes of the growth trajectories were $\beta_0 = 0.1$, $\beta_1 = 0.1$, and $\beta_2 = 0.2$, respectively. The item difficulty parameters were generated from a normal distribution of mean 0 and variance 1, $\delta_i \sim N(0, 1)$ ($i = 1, \ldots, 9$), and $\delta_{10} = -\sum_{i=1}^9 \delta_i$.

In Type I mobility, because students’ school membership changes simultaneously from middle to high school, combinations of middle and high school membership for each student need to be generated. In the present study, a multistage sampling method, which is commonly employed in large-scale assessments, was selected to manipulate combined school membership. For example, in typical educational large-scale surveys, school districts are sampled first; schools from each selected district are sampled next; and then students in every selected school are sampled. In this simulation, it was assumed that there were 10 school districts, and 10 middle schools.
I.-H. Choi and M. Wilson

were fitted for each dataset. In M1, students’ school membership was not considered, and the data structure followed the strict hierarchy. For the CCMM-MLIRD analysis, two different models were specified: one with constant coefficients for the school effects (M2; \( Z_2 = (1, 1, 1, 1)' \) and \( Z_3 = (0, 0, 1, 1, 1)' \)), and the other assuming varied coefficients (M3; \( Z_2 = (1, w_{12}, w_{13}, w_{14}, w_{15})' \) and \( Z_3 = (0, 0, 1, w_{24}, w_{25})' \)). Therefore, M3 was the data-generating model, and the coefficients \( w_{12}, w_{13}, w_{14}, w_{15}, w_{24}, \) and \( w_{25} \) were estimated in this model. A total of 60 replications were made. For each estimator, bias and root mean square error (RMSE) were computed as follows:

\[
\text{Bias}(\hat{\xi}) = \sum_{r=1}^{R} (\hat{\xi}_r - \xi) / R, \quad (15)
\]

\[
\text{RMSE}(\hat{\xi}) = \left( \frac{1}{R} \sum_{r=1}^{R} (\hat{\xi}_r - \xi)^2 / R \right)^{1/2}, \quad (16)
\]

where \( R \) is the number of replications (i.e., 60), and \( \xi \) and \( \hat{\xi}_r \) represent the true value and the parameter estimate from the \( r \)th replication, respectively. According to Hoogland and Boomsma (1998), the estimator was acceptable when the absolute value of relative bias (ARB), which was computed as

\[
\text{ARB}(\hat{\xi}) = \left| \frac{\text{Bias}(\hat{\xi})}{\xi} \right|, \quad (17)
\]

was less than 0.05. In addition to parameter recovery, the accuracy of estimated standard errors was evaluated using the relative bias (RB) of the standard error (Hoogland & Boomsma, 1998),

\[
\text{RB}(\tilde{S}(\hat{\xi})) = (\tilde{S}(\hat{\xi}) - S(\hat{\xi})) / S(\hat{\xi}), \quad (18)
\]

where \( \tilde{S}(\hat{\xi}) \) is the mean of the estimated standard errors across replications, and \( S(\hat{\xi}) \) is the true empirical standard error, which is calculated as the standard deviation of the estimates.

Results

For model selection, the smaller the DIC, the better the fit. According to the guideline suggested by Leckie (2009), the minimum difference of 10 in DIC was used as the cutoff for substantial difference in model fit. Across the 60 replications, the estimated DIC values of M3 were smallest among three models and the differences were substantial, which suggested a 100% correct model detection. It should be noted that the perfect model detection could be due to medium size of variance of student- and school-level random effects and large sample size employed in the simulation study. Specifically, the average of the DIC values of M3 across the replications was 171,326.5, and those of M1 and M2 were 172,050.4 and 171,326.5 respectively.
and the RMSE ranged from 0.011 to 0.088. Even though parameters of the three models are listed in Table 2. The bias of M3 ranged in magnitude from −0.009 to 0.018, and the RMSE ranged from 0.011 to 0.088. Even though the bias values of the $w_{1t}$ and $w_{2t}$ were slightly greater than those of the other estimates, none of the bias estimates was significantly different from zero at the $\alpha = 0.05$ level according to one-sample $t$ tests. The ARB values were acceptable, ranging from 0 to 0.024. These results suggested that the estimates of the data-generating model were unbiased.

Comparisons across the three models revealed that the estimates of the fixed effects including the item difficulty and growth trajectory parameters remained unbiased under M1 and M2 as well. The parameters of the fixed effects were unaffected when the school-level random effects were ignored (M1) and the incorrect coefficients of the school-level random effects were assumed (M2). However, the variance estimates of the random effects were influenced by the misspecification. Under M1, assuming $Z_2 = Z_3 = 0$ resulted in overestimation of the variance of the time-level residual ($\hat{\sigma}_t^2$) and the student-level random effect ($\hat{\psi}_t^2$), yielding ARB($\hat{\sigma}_t^2$) = 0.278 and ARB($\hat{\psi}_t^2$) = 0.746. On the other hand, in M2, the variance of the middle school random effect ($\hat{\tau}_2^2$) was underestimated (ARB($\hat{\tau}_2^2$) = 0.566) and the variance of the high school random effect ($\hat{\tau}_2^2$) was overestimated (ARB($\hat{\tau}_2^2$) = 0.455), while the estimated bias values of $\hat{\sigma}_t^2$ and $\hat{\psi}_t^2$ were acceptable.

Furthermore, although the parameters of the fixed effects were not affected by misspecification, noticeable differences were found in the estimated standard errors of the growth trajectory parameters, $\beta_0$, $\beta_1$, and $\beta_2$, across the three models. The relative biases of the standard errors of $\beta_0$, $\beta_1$, and $\beta_2$ were −0.597, −0.318, and −0.173 in M1, and in M2, they were −0.294, −0.125, and −0.198, whereas in M3 the corresponding values were −0.041, −0.055, and 0.010. In summary, in M1 and M2, the standard errors of the growth trajectory parameters were underestimated, but they were acceptable in M3.

### Type II mobility

#### Data generation

The data with Type II mobility were generated using the CCMM-MLIRD with the linear growth model, Equations (9) and (10). The number of items and measurement occasions were specified as 10 ($I = 10$) and 3 ($T = 3$), respectively. The time variable $d_t$ took on the values of 0, 1, and 2 corresponding to occasions 1, 2, and 3. The Level 2 residual $\epsilon_{ip(j)}$ was generated from a normal distribution with mean 0 and variance 0.4 ($\sigma^2 = 0.4$). The student-specific random effect $\zeta_{ij}$ and the school-specific random effect $\nu_{0k}$ were generated to be normally distributed with mean 0 and variance 0.2 ($\psi^2 = 0.2$ and $\tau^2 = 0.2$). The fixed intercept and slope of the growth trajectories were assumed as $\beta_0 = 0.4$ and $\beta_1 = 0.2$. The same values of item difficulty parameters generated for the Type I simulation were used.

A few conditions related to the cross-classified and multiple membership models have been considered in previous simulation studies; for example, the number of schools and students per school, the magnitude of the variance of the random effects, the intraclass correlations, and the mobility rate. Among these conditions, the mobility rate was the most significant factor influencing the observed bias (Chung & Beretvas, 2011; Grady & Beretvas, 2010; Luo & Kwok, 2012). For such a reason, two conditions of the mobility rate were specified as 10% and 20% in this simulation study. It was assumed that there were 100 schools ($S = 100$) at the first occasion, and 30 students were assigned to each school. Therefore, there were 3,000 students ($J = 3,000$), and a randomly chosen 10% or 20% out of 3,000 students moved to another school between occasions 1 and 2 as well as between occasions 2 and 3.

### Table 2. Bias and root mean square error (RMSE) of the Type I simulation study.

|        | M1 | M2 | M3 |
|--------|----|----|----|
|        | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| $\delta_1$ | 0.002 | 0.019 | 0.002 | 0.019 | 0.002 | 0.019 |
| $\delta_2$ | 0.000 | 0.015 | 0.000 | 0.015 | 0.000 | 0.015 |
| $\delta_3$ | 0.002 | 0.020 | 0.002 | 0.020 | 0.002 | 0.020 |
| $\delta_4$ | −0.001 | 0.021 | 0.000 | 0.021 | 0.000 | 0.021 |
| $\delta_5$ | −0.002 | 0.016 | 0.000 | 0.016 | 0.000 | 0.016 |
| $\delta_6$ | −0.002 | 0.018 | −0.002 | 0.018 | −0.002 | 0.018 |
| $\delta_7$ | 0.002 | 0.015 | 0.001 | 0.015 | 0.001 | 0.015 |
| $\delta_8$ | −0.002 | 0.019 | −0.002 | 0.019 | −0.002 | 0.019 |
| $\delta_9$ | −0.002 | 0.017 | −0.002 | 0.018 | −0.002 | 0.018 |
| $\beta_0$ | 0.009 | 0.05 | 0.007 | 0.049 | 0.007 | 0.050 |
| $\beta_1$ | −0.002 | 0.018 | 0.000 | 0.017 | −0.001 | 0.018 |
| $\beta_2$ | −0.005 | 0.016 | −0.002 | 0.016 | 0.000 | 0.016 |
| $W_{1t}$ | 0.003 | 0.059 | 0.000 | 0.061 | 0.000 | 0.063 |
| $W_{2t}$ | −0.005 | 0.055 | 0.000 | 0.055 | 0.000 | 0.055 |
| $W_{3t}$ | 0.019 | 0.078 | 0.018 | 0.088 | 0.018 | 0.088 |
| $\sigma^2$ | 0.111 | 0.112 | 0.024 | 0.027 | 0.033 | 0.032 |
| $\psi^2$ | 0.149 | 0.151 | −0.006 | 0.012 | −0.001 | 0.011 |
| $\tau_1^2$ | −0.113 | 0.114 | 0.005 | 0.040 | 0.005 | 0.040 |
| $\tau_2^2$ | 0.091 | 0.094 | −0.005 | 0.025 | −0.005 | 0.025 |

Note. Type I simulation study involves Type I school mobility in which students have cross-classified membership of middle and high schools. M1 = three-level HGLM-MLIRD; M2 = CCMM-MLIRD with constant coefficients for the school effects; M3 = CCMM-MLIRD with varied coefficients for the school effects (data-generating model).
**Analysis**

After the data were generated, each data set was analyzed using three models: the three-level HGLM-MLIRD that ignored students’ school membership (M1), the four-level HGLM-MLIRD that assumed that students stayed in the same schools (M2), and the CCMM-MLIRD for Type II mobility used to generate the data (M3). In M1, the strict three-level data structure was assumed, in which responses (Level 1) were nested within measurement occasions (Level 2), and occasions were nested within students (Level 3). On the other hand, in M2, school level (Level 4) was included, but students were assumed to remain within the same school assigned at occasion 1 over repeated occasions. Therefore, some students who switched schools had the wrong school membership at occasions 2 and 3. In M3, students’ correct school membership, which varied across measurement occasions, was considered, and the effects of the multiple schools on students’ growth were investigated. As in the simulation study for Type I mobility, a total of 60 replications were made for each condition of the mobility rate.

**Results**

Similar to the results of the first simulation study, the data-generating model, M3, had the smallest DIC values across the 60 replications under the two mobility rate conditions. Hence, the CCMM-MLIRD was the better-fitting model compared to the three-level and four-level models. Under the 10% mobility rate, the average DIC value of M3 was 101,483.9, and those of M1 and M2 were 101,756.57 and 101,516.9, respectively. In addition, under the 20% mobility rate, the average DIC value of M3 was 101,483.9, and those of M1 and M2 were 101,756.57 and 101,543.1 (M3). In M1, the time-level random effect (\(\hat{\tau}^2\)) was underestimated and the estimated bias increased alongside the mobility rate (ARB(\(\hat{\tau}^2\)) increased from 0.09 (10%) to 0.173 (20%), which was augmented from 10% to 20%. With respect to the standard error estimates, in M1, the standard error of the intercept (\(\hat{\beta}_0\)) was underestimated (RB(S(\(\hat{\beta}_0\))) = −0.570 in the 10% mobility and RB(S(\(\hat{\beta}_0\))) = −0.544 in the 20% mobility), but the standard errors were acceptable in M2 and M3.

**Empirical data study**

As examples of empirical data analysis, data from two large-scale longitudinal assessments were analyzed: the Korean Youth Panel Survey (KYPS) for Type I mobility and the Early Childhood Longitudinal Study–Kindergarten Class (ECLS-K) for Type II mobility. These two data sets were chosen as examples because each type of school mobility was present in these data sets, and previously they have been analyzed using the cross-classified and multiple membership models for cross-sectional or continuous longitudinal outcomes (KYPS by Jeon and Rabe-Hesketh (2012) and ECLS-K by Grady and Beretvas (2010), Luo and Kowk (2012), and Palardy (2010)).

**Example 1: The Korean Youth Panel Survey (KYPS)**

**Data source**

The KYPS, collected by the National Youth Policy Institute (NYPI) in South Korea, was used as an example of Type I mobility. The first survey was administered in 2003 to second-year middle school students, who were followed every year from 2004 to 2007. As mentioned earlier, in the KYPS data, because the students graduated from their middle schools and moved to high schools between the second and third measurement occasions, the data structure involved Type I mobility.

The dependent variables of interest were the responses on 14 items: seven items intended to measure student maturity regarding specific occupation selection (item 1 ~ item 7) and seven items regarding decision related to the students’ future career path in general (item 8 ~ item 14). The contents of the items are given in Appendix. Because the items were stated in a negatively oriented form (e.g., I don’t know well my talents), the 5-point Likert-type responses were dichotomized as following: “strongly disagree” and “disagree” were recorded as 1 and “strongly agree,” “agree,” and “neutral” as 0. To examine the effects of schools on the growth of student maturity in making plans and choosing an occupation, the sample of 2,582 students with full information on school identification (school ID) at each measurement occasion and complete data on the dependent variables were selected for
Before proceeding with the main results, to examine the extent to which the data fit to the Rasch-type modeling (e.g., a constant slope or discrimination parameter across items), item fit was investigated using weighted mean square (infit MNSQ) and the corresponding t statistic. An item was considered as misfit if $|t| > 1.96$ and if infit MNSQ < 0.75 or infit MNSQ > 1.33. In the KYPS, none of the items across multiple timepoints was misfitting. In addition, to evaluate the unidimensionality assumption, we fitted the two-dimensional Rasch model, which assumed two separate dimensions for maturity regarding occupation selection (Item 1 ~ Item 7) and maturity in deciding upon a career path (Item 8 ~ Item 14). Across the five timepoints, the correlations between the two dimensions from the two-dimensional Rasch model were very high (mean of the correlations $= 0.938$), and the two-dimensional model did not improve model fit compared to the unidimensional model. All of these results appeared to support the use of the unidimensional Rasch model for the measurement model at Level 1.

The KYPS data were analyzed using three models: the three-level HGLM-MLIRD (M1), the CCMM-MLIRD for Type I mobility assuming the constant effects of schools over time (M2), and the CCMM-MLIRD for Type I mobility with the varied school effects (M3). The parameter estimates and associated standard errors as well as the deviance, DIC, and PPMC values estimated using the three models are provided in Table 4. The DIC and PPMC values of M3 were the lowest among

**Note.** Type I simulation study involves Type II school mobility in which students have multiple school membership. M1 = three-level HGLM-MLIRD; M2 = four-level HGLM-MLIRD; M3 = CCMM-MLIRD (data-generating model).
the three models. The CCMM-MLIRD, which incorporated students’ Type I mobility and allowed for the varied effects of schools over time, explained the growth of students’ vocational maturity better than the three-level model and the CCMM-MLIRD with constant school effects.

In M3, the difficulties of the first four items were estimated to be positive and the last three items to be negative in both the occupation selection ($\hat{\delta}_1 \sim \hat{\delta}_7$) and career path items ($\hat{\delta}_{8} \sim \hat{\delta}_{14}$). Moreover, the patterns of the estimated item difficulties were similar in the two respects; for example, the students seemed to experience the most difficulty with gaining access to enough occupational and career information ($\hat{\delta}_2 = 1.090$ and $\hat{\delta}_9 = 1.066$), and it was not relatively difficult for them to resolve conflicts with parents ($\hat{\delta}_5 = -1.013$ and $\hat{\delta}_{12} = -1.076$). The fixed regression coefficients of the two-piece linear growth model suggested that student awareness and future planning increased more quickly while attending high school ($\hat{\beta}_2 = 0.247$) than while attending middle school ($\hat{\beta}_1 = 0.130$).

In addition to the item difficulties and growth parameters, the coefficients of the school effects (i.e., factor loadings) were estimated in M3. The coefficient of the middle school effect at occasion 1 was set to one ($w_{11} = 1$); therefore, the estimated coefficients represent how middle schools contribute to the response variable at the current timepoint compared to the initial observation. The estimated coefficients of the middle school effect at occasions 2 and 3 ($\hat{w}_{12} = 0.154$ and $\hat{w}_{13} = 0.346$) suggested a decline of the school effect; however, they did not differ significantly from zero at the $\alpha = 0.05$ level. At occasions 4 and 5, the coefficients were negative ($\hat{w}_{14} = -1.457$ and $\hat{w}_{15} = -1.015$), which meant that the middle schools contributed inversely after the students moved to the high schools, and these estimates were significant at the $\alpha = .05$ level. A positive middle school effect contributes negatively and a negative middle school effect contributes positively to the response variable at occasions 4 and 5. In other words, students tend to move (i.e., regress) toward the mean (i.e., the average student); thus, students with high school effects during occasions 2 and 3 get lower school effects after the students leave middle school (McCaffrey et al., 2004). In contrast to the middle school effect, the coefficients of the high school effect were greater than 1 at occasions 4 and 5 ($\hat{w}_{24} = 1.104$ and $\hat{w}_{25} = 1.826$); therefore, the high school effect was maintained and appeared to increase over time.

The variance estimates of the random effects in M3 suggested that the between-student variance ($\hat{\psi}_2 = 1.071$) and the within-student variance ($\hat{\sigma}^2 = 1.081$) were greater than the between-middle school and the between–high school variances ($\hat{\tau}_1^2 = 0.067$ and $\hat{\tau}_2^2 = 0.111$). Furthermore, there was more variability of the random effects between the high schools than between the middle schools.

As shown in the simulation study, the fixed effect parameters were very similar across the three models. However, in M1, the between-student variance and the within-student variance were estimated to be greater than those in M3. M2 yielded larger estimates of the between-middle school and the between–high school variances compared to M3. Also note that the standard errors of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ in M1 and M2 were less than those in M3.

**Example 2: The Early Childhood Longitudinal Study–Kindergarten Class (ECLS-K)**

**Data source**

A goal of the ECLS-K was to promote the extensive understanding of children’s development from kindergarten to middle school, including the academic performance and social–emotional aspects. To achieve this goal, children who attended kindergarten during the 1998–99 school year were followed through the eighth grade. The data were collected in the fall and spring of kindergarten (1998–99), the fall and spring of first grade (1999–2000), the spring of third grade (2002), the spring of fifth grade (2004), and the spring of eighth grade (2007). The dependent measures of interest in this study were the responses on the math achievement tests at the first three occasions: the spring of kindergarten, first grade, and third grade.

A matrix sampling of items, which is common in large-scale assessments, was adopted in the ECLS-K; thus, each student was administered a particular subset of items. For the purpose of anchoring different test forms across the timepoints and examinees, 14 common items were presented at least on two occasions for the same student from kindergarten to the third grade. In the present analysis, a subset of the responses on these common items was selected as dependent variables. Among the 14 items, the actual number of items that the students responded to ranged between five and 14 on each occasion. A correct response to an item was scored as 1, and an incorrect response as 0. The sample consisted of 4,260 students, and there were 380 schools. The average number of students per school was 10. Of the 4,260 students, 3,910 (91.78%) attended the same school throughout the three measurement occasions; 330 (7.75%) attended two schools; and 20 (0.47%) students attended three schools.

**Results**

Similar to the results for the KYPS data, the analysis of the item fit statistics suggested that the data fitted to the Rasch model well. Moreover, to examine whether the linear growth model was sufficient to explain students’ growth over time, the three-level HGLM-MLIRD with
Table 5. Results from The Early Childhood Longitudinal Study-Kindergarten Classes (ECLS-K) mathematics data analysis.

|       | M1           | M2           | M3           |
|-------|--------------|--------------|--------------|
|       | Est. | SE  | Est. | SE  | Est. | SE  |
| $\delta_1$ | $-1.017$ | $0.027$ | $-1.017$ | $0.027$ | $-1.017$ | $0.026$ |
| $\delta_2$ | $-2.384$ | $0.032$ | $-2.383$ | $0.032$ | $-2.383$ | $0.032$ |
| $\delta_3$ | $-1.898$ | $0.030$ | $-1.898$ | $0.029$ | $-1.898$ | $0.029$ |
| $\delta_4$ | $-1.272$ | $0.027$ | $-1.272$ | $0.027$ | $-1.272$ | $0.027$ |
| $\delta_5$ | $0.319$ | $0.024$ | $0.319$ | $0.024$ | $0.319$ | $0.024$ |
| $\delta_6$ | $0.407$ | $0.024$ | $0.407$ | $0.024$ | $0.407$ | $0.024$ |
| $\delta_7$ | $1.153$ | $0.025$ | $1.153$ | $0.025$ | $1.152$ | $0.025$ |
| $\delta_8$ | $0.624$ | $0.024$ | $0.624$ | $0.024$ | $0.623$ | $0.024$ |
| $\delta_9$ | $-1.639$ | $0.028$ | $-1.640$ | $0.029$ | $-1.639$ | $0.028$ |
| $\delta_{10}$ | $1.044$ | $0.029$ | $1.044$ | $0.029$ | $1.045$ | $0.029$ |
| $\delta_{11}$ | $1.115$ | $0.029$ | $1.115$ | $0.029$ | $1.115$ | $0.029$ |
| $\delta_{12}$ | $1.335$ | $0.030$ | $1.335$ | $0.030$ | $1.336$ | $0.030$ |
| $\delta_{13}$ | $1.809$ | $0.027$ | $1.809$ | $0.027$ | $1.809$ | $0.027$ |
| $\delta_{14}$ | $0.404$ | $0.040$ | $0.404$ | $0.040$ | $0.404$ | $0.040$ |
| $\rho_0$ | $-1.580$ | $0.023$ | $-1.593$ | $0.029$ | $-1.594$ | $0.027$ |
| $\rho_1$ | $2.826$ | $0.031$ | $2.826$ | $0.030$ | $2.825$ | $0.030$ |
| $\rho_2$ | $-0.407$ | $0.009$ | $-0.407$ | $0.009$ | $-0.408$ | $0.009$ |
| $\sigma^2$ | $0.229$ | $0.017$ | $0.229$ | $0.017$ | $0.230$ | $0.017$ |
| $\psi^2$ | $1.091$ | $0.033$ | $0.975$ | $0.032$ | $0.981$ | $0.032$ |
| $\tau^2$ | $0.119$ | $0.018$ | $0.100$ | $0.016$ | $0.100$ | $0.016$ |

Deviance | 115,533.6 | 115,538.8 | 115,537.9 |
DIC | 120,926 | 120,909 | 120,895 |
PPMC | 0.231 | 0.211 | 0.137 |

Note. M1 = three-level HGLM-MLIRD; M2 = four-level HGLM-MLIRD; M3 = CCMM-MLIRD; Est. = estimate; SE = standard error; DIC = deviance information criterion; PPMC = posterior predictive model checking.

Discussion and conclusions

In the present study, the cross-classified and multiple membership models were developed to incorporate students’ school mobility in multilevel and longitudinal item response data. The two types of school switching, specifically, where all of the students switched schools simultaneously at some point (Type I) or where some of the students changed schools at individually varying times during the data collection (Type II), were described, and the corresponding models were proposed. The results of the simulation studies suggested that appropriate modeling of Type I and Type II mobility for school membership using the CCMM-MLIRD yielded good recovery of the parameters. In other words, the proposed models allowed us to estimate item, student, and school parameters simultaneously and to investigate student growth over time, in the presence of the complicated data structures due to students’ school mobility.

Another goal of this study was to investigate the effect of incorrectly modeling school membership. In both of the two types of mobility, the fixed effect estimates were not affected by misspecification of the school-level random effects as shown in previous studies (e.g., Chung & Beretvas, 2011; Grady & Beretvas, 2010; Luo & Kwok, 2012; Meyers & Beretvas, 2006). However, the consequences of ignoring or incorrectly modeling the school-level random effects on the variance estimates of the random effects and the standard errors of the fixed effects were dissimilar according to mobility patterns and model specifications, which are discussed in detail as follows.

First, in Type I mobility, because the time-level residual ($\epsilon_{t,m,h}$), student-level random effect ($\zeta_0$), and school-level random effects ($\gamma_{0,m}$ and $\eta_{0,h}$) were assumed to be independent of each other and normally distributed with a constant variance, the variance of the latent variable $\theta_{t,m,h}$ of the data-generating model, Equation (7), is written as

$$\text{var}(\theta_{t,m,h}) = \psi^2 + w_{1t}^2 + w_{2t}^2 + \tau_{t}^2 + \sigma^2. \quad (8)$$

If school membership was not considered as in M1 of the Type I simulation study by assuming $w_{1t} = w_{2t} = 0$, the variance of $\theta_{t,m,h}$ was estimated as $\hat{\psi}^2 + \hat{\sigma}^2$, and both the between-student and within-student variances were overestimated, compared to the true values. In M2 of the Type I simulation study, because the school effects were assumed constant over time, $w_{1t}$ and $w_{2t}$ took a value of 1 or 0 according to their status at a certain time. For example, at occasions 1 and 2 (attending the middle school), the variance was estimated as $\hat{\psi}^2 + \hat{\tau}_1^2 + \hat{\tau}_2^2 + \hat{\sigma}^2 + \hat{\tau}_{as}^2$ and, after occasion 3 (moving to the high school), estimated as $\hat{\psi}^2 + \hat{\tau}_1^2 + \hat{\tau}_2^2 + \hat{\sigma}^2 + \hat{\tau}_{as}^2$. It should be noted that $w_{1t}$ was assumed to decrease over time and $w_{2t}$ to increase over time in M3. Thus, in M2, $w_{1t}$ was always a linear growth model ($d_t = 0, 1, 3$ for the three time-points) was compared to the three-level models with a quadratic growth model and a cubic growth model. The three-level HGLM-LIRD with a quadratic growth model fitted best among the three models; thus, the quadratic growth model was chosen as the Level 2 model for the three-level HGLM (M1), the four-level HGLM (M2), and the CCMM-MLIRD for Type II mobility (M3).

Table 5 gives a summary of the analysis of the ECLS-K data via the three models. As found in the simulation study, M3 was the best-fitting model according to the estimated DIC and PPMC values. In other words, the CCMM-MLIRD was a more appropriate model when Type II mobility was encountered in the data than the strict HGLM-MLIRD, which ignored students’ school membership (M1) or assumed that students stayed in the same schools over time (M2). Regardless of the differences in the model specification related to the school-level random effect, fitting the three models resulted in similar estimates for the fixed effect parameters. However, the variance estimate of the student-specific random effect associated with the intercept ($\hat{\psi}^2$) was larger in M1 than the estimates using M2 and M3. In M2, $\psi^2$ was almost identical to the estimate in M3, but the between-school variance ($\tau^2$) was smaller than in M3. As a whole, these results were consistent with the findings of the simulation study for Type II mobility.
greater than or equal to, and $w_{2t}$ was smaller than or equal to, the corresponding coefficients in M3 (true values). As a consequence of misspecifying the design matrix of the school-level random effects, in M2, the variance of the middle school random effect was underestimated (associated with larger coefficients than the true values). The variance of the high school random effect was also overestimated (associated with smaller coefficients than the true values). Moreover, in M1 and M2, these incorrectly estimated variances of the random effects resulted in underestimation of the standard errors of the intercept and slopes of the growth curve model.

Similarly, in Type II mobility, the variance of the latent variable $\theta_{tj(k)}$ of the true model, Equation (10), is expressed as

$$\text{var}(\theta_{tj(k)}) = \psi^2 + \tau^2 \sum_{k \in (j)} \lambda_{tjk}^2 + \sigma^2. \quad (18)$$

In M1, $\lambda_{tjk}$ was assumed to be zero; thus, the variance was estimated as $\hat{\psi}^2 + \hat{\sigma}^2$, and only the between-student variance is overestimated relative to the true value, yielding underestimation of the standard errors of the intercept of the growth curve model. Under M2, the students were assumed to stay within the first school they attended; thus, $\lambda_{tjk}$ took a value of 1 associated with a school $k$ that student $j$ attended at occasion 1, and the estimated variance was $\hat{\psi}^2 + \hat{\tau}^2 + \hat{\sigma}^2$ at any timepoint $t$. When students’ school mobility was modeled, because $\lambda_{tjk}$ represented the relative contribution of school $k$ on student $j$ at occasion $t$, $\lambda_{tjk}$ was less than or equal to 1. Hence, with the existence of mobile students attending multiple schools, $\sum_{k \in (j)} \lambda_{tjk}^2$ in the true model was always less than or equal to 1, yielding underestimation of the between-school variance in M2 (associated with larger coefficients than the true values). For the same reason, when there are more mobile students—that is, the mobility rate increases—the degree of underestimation of the between-school variance in M2 increases as shown in the present and previous simulation results (e.g., Chung & Beretvas, 2011; Leckie, 2009; Luo & Kwok, 2012).

These results suggest that when the school-level random effects were not included in the models, as in the three-level models, the between-school variance was redistributed to the lower levels. In addition, the use of incorrect design matrices associated with the school-level random effects produced overestimated or underestimated between-school variances. In particular, misspecification of Type I mobility was related to the assumption regarding whether the school effects remained constant, and the consequences of misspecification were determined by the degree to which the school-level random effects changed over time. In the empirical illustration using the KYPS, the CCMM-MLIRD provided a significantly better fit to the data, and the estimated coefficients of the school-level random effects implied that the effects of the schools varied over time. In Type II mobility, the proportion of mobile students determined the effects of misspecification; thus, it is necessary to examine the mobility rate in the data prior to the data analysis. If the mobility rate is not substantial, using only the first school information may not produce very different results from the multiple membership models. However, the results from the simulation study showed that even under the 10% mobility condition, the first school approach yielded underestimation of the variance of the school-level random effect (and, recall, an average mobility rate of 17% was reported; U.S. Government Accounting Office, 1994). Moreover, incorrectly modeling school-level random effects resulted in underestimation of the standard errors of the fixed intercept and/or slopes of the growth curve model, yielding inflated Type I error rates. In summary, ignoring or incorrectly modeling the school-level random effects in analyzing complicated longitudinal item response data could lead researchers to conclude that more or less variability exists than really does. The identification of a substantial variance of the school-level random effects often directs researchers to investigate school characteristics that may explain the variability across schools (Meyers & Beretvas, 2006); therefore, it is important to model the cross-classified and multiple school membership appropriately.

Because this study was a preliminary investigation of how to incorporate mobility into multilevel and longitudinal item response data and to assess the effect of incorrectly modeling group membership, there are several limitations. First, the current article focused on students’ school switch, but it would be important to examine whether the model framework and findings from this study can be applied to multilevel and longitudinal item response data in other disciplines that involve group membership change. Second, although the parameter values in the simulation study were chosen according to the values used in the previous literature related to multilevel modeling and cross-classified and multiple membership models (e.g., Luo & Kwok, 2012; Meyers & Beretvas, 2006; Raudenbush & Liu, 2001), the conditions employed in the study are limited. Broader investigations incorporating factors that reflect the complexity of real data need to be conducted.

For example, in both types of simulation studies, the correct model detection was perfect (100%); however, these results could be due to limitations in the simulation conditions (e.g., medium size of variance of student-and school-level random effects and relatively large sample sizes). In the Type II simulation study, students...
who switched schools between timepoints were randomly selected, but the action of switching schools could be associated with student background and school characteristics. One possible factor is student achievement, and previous studies have shown a negative relation between school change and academic achievement (e.g., Heinlein & Shinn, 2000; Rumberger, 2003; Rumberger & Larson, 1998; Temple & Reynolds, 2000). In the current study, we did not assume any relation between mobility and missing data; however, it is also possible that mobile students are more likely to leave items unanswered, yielding missingness into the data. Moreover, other types of misspecification can be incorporated. As another example, in the Type II simulation, instead of the first school the student attended, the last school information can be included for M2. When only the last school is considered, similarly as for the first school approach, the effects of previous schools, if students attended multiple schools, are not modeled properly. This study focused on large-scale longitudinal assessments and employed relatively large sample size for the simulations and empirical illustrations. Future studies could investigate the performance of the CCMM-LIRD and the effect of misspecification using a smaller sample size. To be specific, in Type I mobility, the standard errors of factor loadings $w_1$ and $w_2$, were estimated to be greater in the real data analysis than in the simulation study. Given that the standard errors of factor loadings are related to sample size (MacCallum, Widaman, Zhang, & Hong, 1999), and those larger standard errors could be due to the smaller number of students within each school in the KYPS data than in the simulated data.

There are also possible extensions of the current framework of the CCMM-MLIRD. Here, we considered only the Rasch versions of the measurement models, but they can be extended to more complex models such as 2PL and 3PL IRT models for binary outcomes. In addition, one could investigate polytomous data using the partial credit model (PCM; Masters, 1982) and the generalized partial credit model (GPCM; Muraki, 1992). In this study, item difficulties are assumed constant over time (e.g., measurement invariance); however, they can also be modeled by including time-variant difficulty parameters in the measurement model. For example, Kelcey et al., (2014) extended cross-classified IRT models for polytomous responses and measurement noninvariance across different raters. Furthermore, student-specific random slopes and student- and school-level explanatory variables can be added to the growth curve model.

Finally, the current study employed Bayesian estimation using MCMC to estimate model parameters. Because the MCMC procedures implemented in WinBUGS required substantial computing time for convergence, which is not uncommon in MCMC estimation, only 60 replications were made. To enhance the practical use of the proposed model, other software that handles the cross-classified and multiple membership models for discrete longitudinal data might be considered for future studies. For example, Mlwin (Rasbash, Steele, Browne, & Goldstein, 2012) and Mplus version 7 (Muthén & Muthén, 1998–2012) allow MCMC estimation for cross-classified models including categorical and continuous dependent variables. Comparisons between WinBUGS and these programs could be further addressed in future simulation studies to extend this current line of research.

**Article information**

**Conflict of Interest Disclosures:** Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

**Ethical Principles:** The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

**Funding:** This work was not supported.

**Role of the Funders/Sponsors:** None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

**Acknowledgements:** The authors thank the two anonymous reviewers and the editor for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors’ institutions is not intended and should not be inferred.

**References**

Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, 22(1), 47–76. doi:10.3102/10769986022001047

Asparouhov, T., & Muthén, B. (2012). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, British Columbia.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, 39(4), 331–348. doi:10.1111/j.1745-3984.2002.tb01146.x

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, 1(2), 103–124. doi:10.1177/1471082X0100100202

Cho, S.-J., & Cohen, A. S. (2010). A multilevel mixture IRT model with an application to DIF. *Journal of...*
Chung, H., & Beretvas, S. N. (2011). The impact of ignoring multiple membership data structures in multilevel models. *British Journal of Mathematical and Statistical Psychology*, 65(2), 185–200. doi:10.1111/j.2044-8317.2011.02023.x

Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42(2), 133–148. doi:10.1111/j.1745-3984.2005.00007

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). An introduction to latent variable growth curve modeling: Concepts, issues, and applications. Mahwah, NJ: Erlbaum.

Gelman, A., & Rubin, D. B. (2004). Bayesian data analysis (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London, England: Anorld.

Grady, M. W., & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research*, 45(3), 393–419. doi:10.1080/002731712.10.483390

Heinlein, L. M., & Shinn, M. (2000). School mobility and student achievement in an urban setting. *Psychology in the Schools*, 37(4), 349–357. doi:10.1002(1520-6807(200007)37:4<359::AID-PITS6>3.0.CO;2-1

Hill, P. W., & Goldstein, H. (1998). Multilevel modeling of educational data with cross-classification and missing identification for units. *Journal of Educational and Behavioral Statistics*, 23(2), 117–128. doi:10.1207/s15327906jabs2302_11

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26(3), 329–367. doi:10.1177/004912419802603003

Hung, L. F., & Wang, W.-C. (2012). The generalized multilevel facets model for longitudinal data. *Journal of Educational and Behavioral Statistics*, 37(2), 231–255. doi:10.3102/1076998611402503

Jeon, M., & Rabe-Hesketh, S. (2012). Profile-likelihood approach for estimating generalized linear mixed models with factor structures. *Journal of Educational and Behavioral Statistics*, 37(4), 518–542. doi:10.3102/1076998611417628

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38(1), 79–93. doi:10.2307/1435439

Kelcey, B., McGinn, D., & Hill, H. (2014). Approximate measurement invariance in cross-classified rater-mediated assessments. [Methods]. *Frontiers in Psychology*, 5, 1469. doi:10.3389/fpsyg.2014.01469

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A*, 172(3), 537–554. doi:10.1111/j.1467-985X.2008.00577.x

Li, F., Duncan, T. E., Duncan, S. C., & Hops, H. (2001). Piecewise growth mixture modeling of adolescent alcohol use data. *Structural Equation Modeling*, 8(2), 175–204. doi:10.1207/S15328007SEM0802_2

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models* (vol. 840). Cary, NC: SAS Institute.

Lockwood, J. R., McCaffrey, D. F., Mariano, L. T., & Setodji, C. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics*, 32(2), 125–150. doi:10.3102/1076998606298039

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility*. Statistics and Computing, 10(4), 325–337. doi:10.1023/A:1008929526011

Luo, W., & Kwok, O.-M. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*, 44(2), 182–212. doi:10.1080/00273170902794214

Luo, W., & Kwok, O.-M. (2012). The consequences of ignoring individuals’ mobility in multilevel growth models: A Monte Carlo study. *Journal of Educational and Behavioral Statistics*, 37(1), 31–56. doi:10.3102/1076998610394366

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84–99. doi:10.1037/1082-989X.A.1.84

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. doi:10.1007/BF02296272

McArdle, J. J. (1988). Dynamic but structural equation models of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (Vol. 2, pp. 561–614). New York, NY: Plenum Press.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58(1), 110–133. doi:10.3102/10769986029001067

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67–101. doi:10.3102/10769986029001067

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, 41(4), 473–497. doi:10.1207/s15327906mbfr4104_3

Mislevy, R. J., & Bock, R. D. (1989). A hierarchical item-response model for educational testing. In R. Bock (Ed.), *Multilevel analysis of educational data* (pp. 57–74). San Diego, CA: Springer.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–176. doi:10.1002/j.2333-8504.1992.tb01436.x

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Youth Policy Institute. (2009). *Korean Youth Panel Survey (KYPs) user's guide for the 1st-5th year of panel study of the second year middle school students*. Seoul, Korea: National Youth Policy Institute.

Palardy, G. J. (2010). The multilevel crossed random effects growth model for estimating teacher and school effects: Issues and extensions. *Educational and Psychological Measurement*, 70(3), 401–419. doi:10.1177/0013164409355693
Appendix. Contents of Items Related to Vocational Maturity of the Korean Youth Panel Survey (KYPS)

| Item No. | Contents |
|----------|----------|
| Specific Occupation Selection | Item 1 | I do not have enough information about my competencies or personal traits to be able to make a specific occupation selection. |
| | Item 2 | I do not have enough information about the kinds and characteristics of various occupations. |
| | Item 3 | I am attracted by more than one occupation and it is difficult to choose one among them. |
| | Item 4 | I constantly make a change in occupation selection. |
| | Item 5 | I find it difficult to make a career decision because of the difference between my parents and myself regarding occupation selection. |
| | Item 6 | It is meaningless to make an occupation selection now because I do not know what the future will be like. |
| | Item 7 | In making an occupation selection, I tend to follow my parents’ recommendations more than my own opinion. |
| General Career Path | Item 8 | I do not have enough information about my competencies or personal traits to be able to make a career path. |
| | Item 9 | I do not have enough information about the kinds and characteristics of various career paths. |
| | Item 10 | I am attracted by a number of career paths, so it is difficult for me to choose one among them. |
| | Item 11 | I constantly make a change in career planning. |
| | Item 12 | I find it difficult to make a career plan because of the differences between my parents and myself regarding career planning. |
| | Item 13 | It is meaningless to set a career plan now because I do not know what the future will be like. |
| | Item 14 | In setting a career plan, I tend to follow my parents’ recommendations more than my own opinion. |

Note. The items were originally administered in Korean; they were translated into English and provided by the National Youth Policy Institute (National Youth Policy Institute, 2009).