Journal of Modern Applied Statistical Methods

Volume 10 | Issue 1 Article 21

5-1-2011

General Piecewise Growth Mixture Model: Word Recognition Development for Different Learners in Different Phases

Amery D. Wu
University of British Columbia, ameryw@yahoo.com

Bruno D. Zumbo
University of British Columbia, bruno.zumbo@ubc.ca

Linda S. Siegel
University of British Columbia

Recommended Citation

Wu, Amery D.; Zumbo, Bruno D.; and Siegel, Linda S. (2011) "General Piecewise Growth Mixture Model: Word Recognition Development for Different Learners in Different Phases," *Journal of Modern Applied Statistical Methods*: Vol. 10 : Iss. 1 , Article 21.
DOI: 10.22237/jmasm/1304223600

Journal of Modern Applied Statistical Methods
May 2011, Vol. 10, No. 1, 226-248

1538 – 9472/11/$95.00

General Piecewise Growth Mixture Model: Word Recognition Development for Different Learners in Different Phases

Amery D. Wu Bruno D. Zumbo Linda S. Siegel
University of British Columbia,
Vancouver, B.C. Canada

The General Piecewise Growth Mixture Model (GPGMM), without losing generality to other fields of study, can answer six crucial research questions regarding children's word recognition development. Using child word recognition data as an example, this study demonstrates the flexibility and versatility of the GPGMM in investigating growth trajectories that are potentially phasic and heterogeneous. The strengths and limitations of the GPGMM and lessons learned from this hands-on experience are discussed.

Key words: Structural equation model, piecewise regression, growth and change, growth mixture model, latent class analysis, population heterogeneity, word recognition, reading development, trajectories, literacy development.

Introduction

People learn and develop in different ways in different phases. A rich body of literature has documented the complexities in human development, among which the best known is probably Piaget's phasic theory about children's cognitive development. However, in statistical modeling, such complexities are often disguised by a primitive assumption about homogeneity and linearity of data. The purpose of this study is, in the context of children's reading development, to demonstrate the application of the General Piecewise Growth Mixture Model (GPGMM). GPGMM is a versatile modeling strategy that allows for the investigation of trajectories that are heterogeneous and phasic. GPGMM marries the general growth mixture model (GGMM) articulated by Muthén (2004) with piecewise regression (Li, Duncan, Duncan, & Hops, 2001; McGee & Carleton, 1970; Muthén & Muthén, 1998-2007).

Overview of Two Reading Development Theories

The debate over the developmental pathways of children's literacy achievement has not been resolved. Two major competing theories exist: the deficit and the lag models. The deficit model suggests that children who have a superior start in precursor linguistic and cognitive skills will improve their reading skills at a faster rate than those with a slower start (e.g., Bast & Reitsma, 1998; Francis, et al., 1996). The increasing difference in reading performance among poor, average and advanced readers observed in early development is believed to be a result of initial skill sets that never develop sufficiently in those who turn out to be poor readers.

An alternative view, the lag model, suggests that children with a poorer start in their cognitive skills will display a faster growth in their later development, whereas those with a superior start will display a slower growth (Leppänen, et al., 2004; Phillips, et al., 2002).

Bruno D. Zumbo is a Professor, Amery D. Wu is a Post-doctoral Research Scientist, and Linda S. Siegel is a Professor at the University of British Columbia. Send correspondence to Bruno D. Zumbo at bruno.zumbo@ubc.ca or Amery Wu at ameryw@yahoo.com.

Protagonists of this view believe that children who differ in reading ability vary only in the rate at which cognitive skills develop so that lagging children will eventually catch up with their peers, and that the gap in the early development will eventually disappear.

Empirical evidence has not consistently confirmed either the deficit or lag model. Bast and Reitsma (1998) provided support for the deficit model based on the findings that the rank ordering of the six waves of word recognition scores remained stable and that the differences in the score increased from grade one to grade three. They concluded that differences in reading achievement of the 280 Dutch children were cumulative. In a longitudinal study, Francis, et al. (1996) studied the trajectories of 403 non-disabled and disabled children in Connecticut from grade one to grade nine using the Rasch-scaled composite score of the Word Identification, Word Attack and Passage Comprehension subtests (Woodcock-Johnson Psychoeducational Test Battery; Woodcock & Johnson, 1977). They used quadratic trajectories to model the non-linear growth pattern displayed in the data. The results showed that the disabled readers were unable to develop adequate reading skills and their problems persisted into adolescence. They concluded that a deficit model best characterized the enlarging gap and an intervention at an early age is essential in order to reduce the impact of early deficit.

Other studies, however, have reported that initially poor readers improved faster, and the early gap decreased over time (e.g., Anrnoutse, et al., 2001; Aunola, et al., 2002; Jordan, Kaplan & Hanich, 2002; Scarborough & Parker, 2003). For example, assuming linearity from grade two to grade eight, Scarborough and Parker (2003) reported decreasing gaps of 57 non-disabled and disabled children in both WJ-Word Identification and WJ-Passage Comprehension.

In a longitudinal study of 198 English readers in Canada from grade one to grade six, Parrila, et al. (2005) studied the development of word identification, word attack and passage comprehension separately. For each outcome measure, they fitted a latent growth quadratic curve using growth mixture modeling and found that children with lower starting performance reduced the distance between themselves and children who had higher initial performance. Aarnoutse, et al. (2001) also failed to find the fan-spread pattern in reading comprehension, vocabulary, spelling or word decoding efficiency. Their results suggested that the initially low performers tended to show greater gains than did medium or high performers. Similarly, Aunola, et al. (2002) found a decrease in individual differences in a reading skill score (a composite of four different reading tasks) of Finnish children. Scarborough and Parker (2003) also reported that the difference between good and poor readers in their US sample were smaller in grade eight than grade two in a composite reading score made of word reading, decoding and passage comprehension.

Existing evidence has not provided conclusive support for either the deficit or the lag models, or for the relationship between early performance and subsequent growth rate. The incongruence in the empirical findings is palpable if careful attention is paid to the diversified and piecemeal approach to the research design and data analysis (Parrila, et al., 2005).

As is evident from this brief review, the research designs varied in the length and phase of the studied time interval (i.e., earlier or later development in the grade school), the statistical analyses (e.g., ANOVA, regression or latent growth model), measures used to represent reading ability, the population of children whose growth trajectories were compared (e.g., normative or children with learning difficulties), the hypothesized pattern of growth trajectory (e.g., linear or quadratic), outcome measure (e.g., word recognition or reading comprehension) and sample size, as well as the terminologies and their operational definitions. Parrila, et al. (2005) concluded that reading development could follow multiple pathways, only some of which are captured by the existing conceptualizations. Thus, researchers could benefit from a more integral and comprehensive data analytical framework that is capable of modeling the complex, intricate, and diversified developmental nature of children's reading development.

Methodology

Data

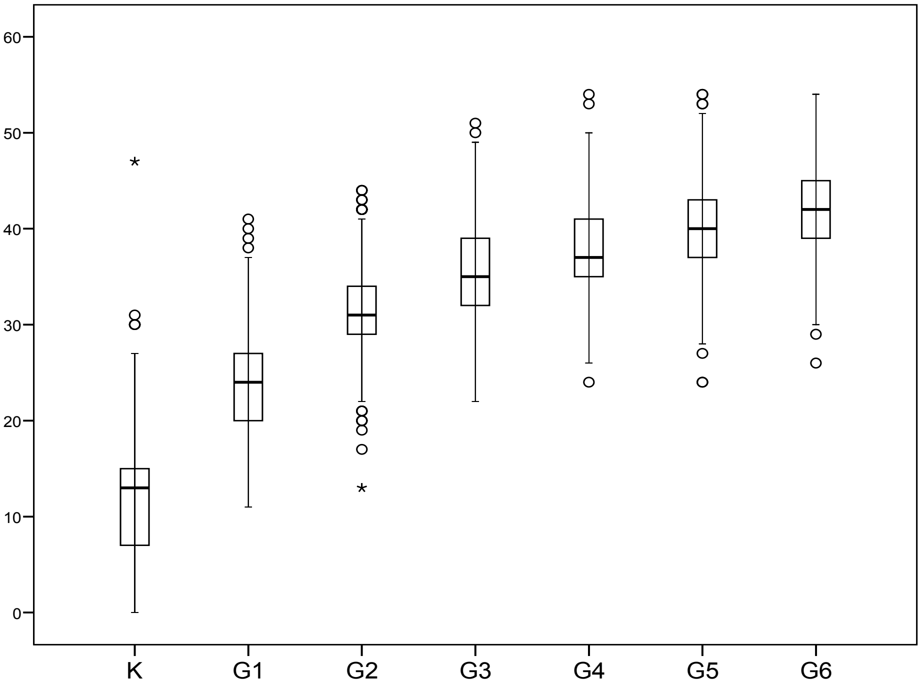
The data consists of 1,853 elementary school children from the North Vancouver school district in British Columbia. These children were measured every year in the fall starting from kindergarten to grade six. The dependent variable, word recognition, had a maximum score of 57, which was measured by naming 15 alphabet letters and followed by the reading subtest of the Wide Range Achievement Test-3 (WRAT-3; Wilkinson, 1995), which has a list of 42 words ordered by difficulty. The measurement of word recognition remained the same across the seven waves of data collection; hence, measurement invariance that warranted temporal score comparability across grades was assumed. For demonstrative purposes, only the data of the 526 children who had all seven waves of data were included. Outliers were retained because this study aimed to model these cases through distinct latent classes so that – within each class – the distribution of reading performance was assumed to be normal.

Table 1 displays the mean (M), standard deviation (SD) and skewness of the seven waves of the data. Figure 1 displays the boxplots for the seven waves of word recognition scores. It can be observed that the distributions of the seven word recognition measures are, for the most part, symmetric. The overall performance in word recognition improved across time, with faster growth in the period between kindergarten and grade two, and relatively slower growth in

Table 1: Descriptive Statistics of the Word Recognition Scores

	K	G1	G2	G3	G4	G5	G6
Mean	11.60	23.86	31.55	35.63	37.62	40.27	41.89
SD	5.14	4.85	4.52	4.99	4.39	4.71	4.20
Skewness	-0.21	0.38	0.22	0.32	0.15	0.07	-0.14

Figure 1: Boxplots for the Word Recognition Scores from Kindergarten to Grade Six



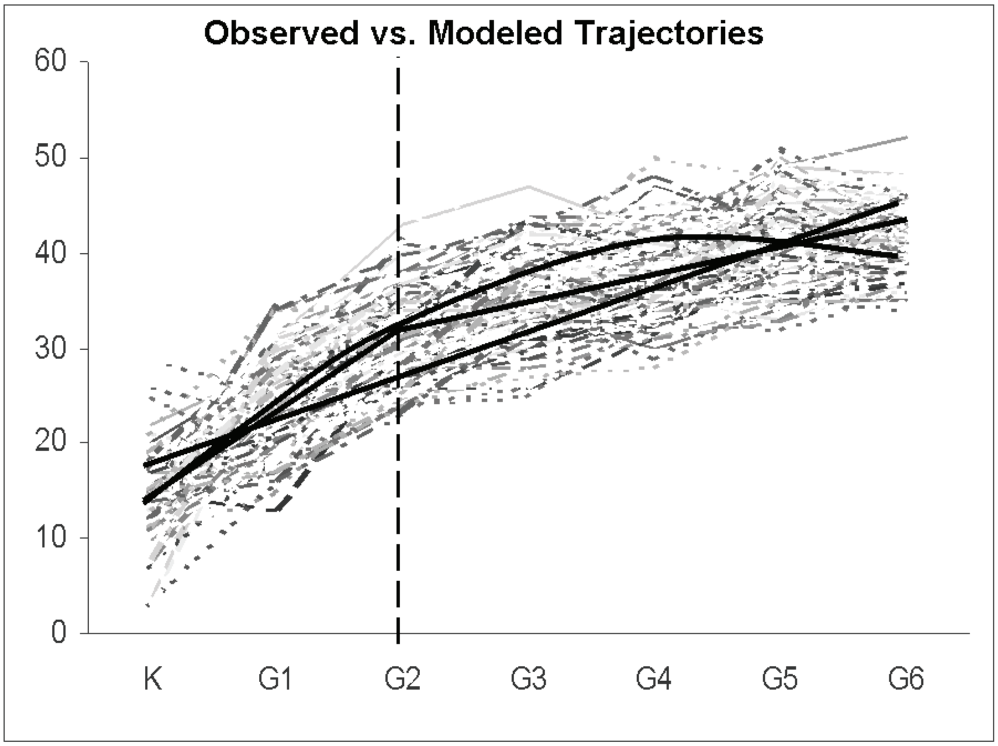
the period between grade three and grade six. Figure 2 shows the individual trajectories. The overall pattern of the trajectories was consistent with those revealed in the boxplots and the literature, which showed a nonlinear trend (Francis, et al., 1996; Parrila, et al., 2005).

Given the observed non-linearity, it would seem inappropriate to impose a linear trajectory to the observed data as portrayed by the thick single straight line in Figure 2. Most previous studies fitted a quadratic curve to model this non-linear pattern as portrayed by the thick curve line in Figure 2, where the early development is assumed to improve with a faster growth, followed by a relatively slower growth, and then reach a peak with a possibility to decline near the end. Although a quadratic function is fairly accessible and widely used by applied researchers, it may be inappropriate for literacy development of school-age children, because it may portray a decline at the end of the developmental course, whereas reading development, at worst, is expected to plateau rather than decline, if not continue to grow. Also, the meaning of a quadratic parameter is often hard to interpret conceptually for phenomena studied in the social and behavioral sciences, such as word recognition.

Another possibility for modeling the developmental pattern observed in Figure 2 is to fit a piecewise linear trajectory (Khoo, 1997; Li, et al., 2001; McGee & Carleton, 1970; Raudenbush & Bryk, 2002) as shown by the two thick segments connected at grade two in Figure 2. A piecewise trajectory allows different linear growth rates to be fitted to different developmental phases that are empirically observed or theoretically hypothesized.

Notice that despite the overall trend observed in Figure 2, a great deal of variation exists in individual's developmental pattern as demonstrated by the differences in the starting performance, the speed of learning over time and the ending performance at grade six. Imposing a homogeneous trajectory to these heterogeneous learning patterns may overlook the complexities and diversity of children's reading development.

Figure 2: Observed vs. Modeled Trajectories (Single Linear, Quadratic and 2-piece Linear) of Word Recognition Scores from Kindergarten to Grade Six



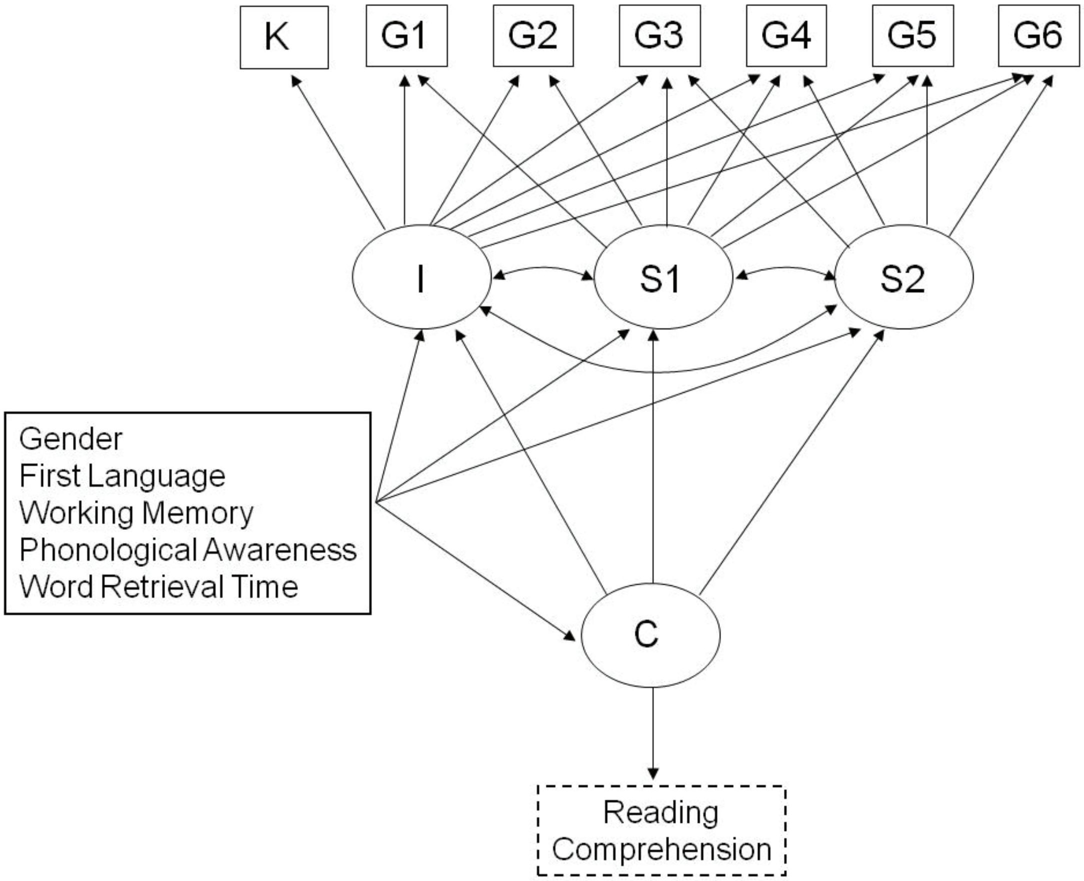
General Piecewise Growth Mixture Model (GPGMM): What Can It Do?

The GPGMM, at its foundation, is a structural equation model, a latent variable approach for investigating growth and change (Meredith & Tisak, 1990; Muthén, 2001; Muthén, 2008). GPGMM is a relatively new and fairly complex modeling framework for studying growth and change (Muthén, 2004). It combines the growth mixture model (GMM) that models population growth heterogeneity with the piecewise regression that models phasic growth rates.

The "mixture" of growth mixture modeling refers to the finite mixture modeling element; that is, modeling with categorical latent variables that represent subpopulations (classes) where population membership is unknown but is inferred from the data (McLachlan & Peel, 2000). The "piecewise" of the piecewise regression refers to the growth rates in different developmental phases as reflected via the continuous latent growth factors. The GPGMM is an extension of the piecewise GMM by adding the covariates and the developmental outcome variable (Muthén, 2004). The combination of continuous and categorical latent variables of the GPGMM provides a very flexible analytical framework for investigating subpopulations showing distinct and phasic developmental patterns.

GMM has gained increased popularity in studying children's reading development. Statisticians and methodologists have proposed growth mixture models other than the GPGMM demonstrated in this paper that are of great theoretical and practical significance. Examples of these developments include Muthén, et al. (2003) and Boscardin, et al. (2008). The GPGMM specified for this demonstration is depicted graphically in Figure 3.

Figure 3: GPGMM for Word Recognition Development



GPGMM attends to individual differences in developmental changes by allowing the growth factor to vary across individuals, resulting in individual varying trajectories over time. This individual heterogeneity in trajectories, in a conventional linear form, is captured by two continuous growth factors (a.k.a., random effects); one is a latent variable representing individual differences in the initial performance (i.e., intercept), and the other representing the individual differences in the growth rate (i.e., slope). Growth factors are created by summarizing the growth patterns observed in the repeated measures of the same individuals over time. The categorical latent variable C in Figure 3 models the population heterogeneity in the growth factors.

GPGMM can answer six crucial questions pertaining to children's reading developmental complexities. These research questions, shown in two sequential sets, are:

Set A:

A1. Are there distinct phases where children differ in their speed of learning?

A2. Are there unknown subpopulations (latent classes) that differ in their growth pattern?

A3. How are the starting performance and growth rates related?

Set B:

B1. What are the characteristics of the latent classes?

B2. For each class, what explains children's starting performance and growth rates?

B3. Do the latent classes differ in the reading developmental outcome?

Note that although these questions were posed and answered herein as two sequential sets (A1-A3 and B1-B3), the experience with the modeling procedures of the GPGMM in this study was non-linear; it required a recursive process of model specification, model selection, and meaning checking as would be the case with any other complex modeling. Nonetheless, two general modeling stages can be distinguished for a GPGMM. The first stage was the unconditional piecewise GMM (i.e., without the covariates and the proximal developmental outcome variable), that answered questions A1, A2 and A3. The major modeling task of this stage was to choose the optimal growth trajectories and number of classes. The second stage entailed the full GPGMM by incorporating the covariates and the developmental outcome variable (i.e., reading comprehension measured at the last time point) into the unconditional piecewise GMM. The conditional piecewise GMM further answered questions B1, B2, and B3. The major modeling task of this stage was to understand the characteristic of the classes and explain class-specific variations in the growth factors and the developmental outcome variable.

Model Estimation and Fit

The following briefly describes the model estimation method and fit statistics used for this demonstration. As asserted at the outset, the focus of this article is to provide a conceptual account and modeling demonstration of the GPGMM instead of technical details. General model specification can be found in Technical Appendix 8 of M*plus* (Muthén, 1998-2004) and M*plus* User's Guide (Muthén & Muthén, 1998-2007). The technical details can be found in Muthén and Shedden (1999) and Muthén and Asparouhov (2008). To foster a wider use of the GPGMM, the M*plus* syntax for the final model can be found in the Appendix.

In M*plus*, three estimators are available for a GMM: (1) maximum likelihood parameter estimates with conventional standard errors (ML), (2) maximum likelihood parameter estimates with standard errors approximated by first-order derivatives (MLF), and (3) maximum likelihood parameter estimates with robust standard errors (MLR). The major difference among these estimators lies in the approach for approximating the Fisher information matrix.

The MLR is designed to be robust against non-normality and misspecification of the likelihood. Simulation studies have suggested that MLR standard errors perform slightly better than those of ML, and the standard errors of ML perform better than those of MLF (for details see Technical Appendix 8 of M*plus*; Muthén & Muthén, 1998-2007). In this application, the GPGMM parameters were estimated in M*plus* 5.21 with the default MLR estimation, since it is designed to model the potential population non-normality due to the potentially unknown subpopulations. We also adopted the default number of 15 numerical integration points (Muthén & Muthén, 1998-2007), because increasing the integration points can substantially increase the time for estimating a complex model like the GPGMM.

When a mixture model is specified, M*plus* uses random starts to guard against local maxima. The default starting values are automatically generated values that are used to create randomly perturbed sets of starting values for all parameters in the model except variances and covariances. Throughout the analyses, the number of initial stage random starts were, as a principle, first set to 1,000, and the final stage starts were set to 20 (e.g., the syntax reads STARTS = 1,000 20). If the log-likelihood values were not replicated as reported in the final 20 solutions, the number of the initial random starts was increased until the log-likelihood was replicated at least twice. For all analyses, the initial stage iterations are set to 200 and the maximum number of iterations for the EM algorithm was set to 3,000.

To speed up the estimation, M*plus* allows user-specified starting values. In this application, four strategies were considered for specifying the starting values. The first and simplest strategy was to specify some or all of the starting values to zeros; this would significantly reduce the computing time. The second strategy was to use descriptive statistics obtained from the given data or reported in the literature (e.g., the mean of the WRAT-3 at the kindergarten year as the starting value of the intercept growth factor. The third strategy was to estimate a multi-class model with the variances and covariances of the growth factors fixed at zero. The estimates of the growth factor means from this analysis were then used as the starting values in the analysis where the growth factor variances and covariances were freely estimated. The fourth strategy was to use the estimates from a simpler model as the starting values for a more complex model. For example, the estimated means of the growth factors from the 1-class model were used for the 2-class model or the growth factor means estimated from the unconditional piecewise GMM were used as the starting values for the conditional piecewise GMM. These methods for specifying starting values were used interchangeably and in concert to help the model estimation.

In the demonstration, the quality of a GMM model was assessed by several fit statistics and two alternative likelihood ratio tests (LRT). The conventional test of model fit based on the Chi-square likelihood ratio, comparing a compact model (K-1 classes) with an augmented model (K classes), does not function properly because it does not have the usual large-sample chi-square distribution. Two alternative likelihood-based tests have been developed to overcome this problem and have shown promise.

The first is the Lo, Mendell, and Rubin (2001) likelihood ratio test (LMR LRT; Lo, Mendell & Rubin, 2001; Nylund, Asparouhov, & Muthén, 2007). Assuming within class normality, this test proposes an approximation to the conventional distribution of likelihood ratio test and provides a *p*-value for testing K-1 classes against K class. A low *p*-value indicates that a K−1 class model has to be rejected in favor of a model with at least K classes. The second was the bootstrapped parametric likelihood ratio test (BLRT, described in McLachlan & Peel, 2000). As opposed to assuming that twice the difference between the two negative log-likelihoods follows a known distribution, the BLRT bootstraps samples to estimate the difference distribution based on the given data. The interpretation of the BLRT *p*-value is similar to that of the LMR LRT. Both LMR LRT and BLRT are available in M*plus* in the Technical Output 11 and 14 respectively.

Another type of commonly used fit indices is the information criterion: Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Sample-Size Adjusted BIC (SBIC). These fit indices are scaled so that a small value corresponds to a better model with a large log-likelihood value and not too many parameters. The SBIC was found to give superior performance in a simulation study for latent class models in Yang (2006), and the BIC was found to give superior performance for mixture models including the GMM (Nylund, et al., 2007). Note that these indices do not address how well the model fit to the data, but are relative fit measures comparing competing models.

When classification is a major modeling concern as with a GMM, the classifying quality is often assessed. Entropy assesses the degree to which the latent classes are clearly distinguishable by the data and the model. It is scaled to have a maximum value of 1 with a high value indicating a better classification quality. Entropy is calculated based on individual's estimated posterior probabilities of being in each of the K classes (analogous to factor scores in a factor analysis).

Consider each individual is classified into one of the K classes by the highest estimated posterior probability (i.e., most likely classes), entropy value will approach one if individuals' probabilities in the K classes approach one or zero, whereas the entropy value decreases if individuals' posterior probabilities of being in the K classes depart from zero or one (see Technical Appendix 8 for the calculation of Entropy (Muthén, 1998-2007; Clark & Muthén, 2009). In other words, Entropy reflects how much noise there is in the classification, hence, can be understood as an index for classification reliability.

There is not yet consensus upon the level of satisfactory entropy. Clark and Muthén (2009), in studying the effect of entropy on relating the latent classes to covariates, arbitrarily used the value of 0.8 as an indication of high entropy; thus, this is the minimum value that was used 0.8 for being considered as reliable class classification. All the aforementioned fit indices and LRTs were reported and examined in concert to choose an optimal number of classes.

Unconditional Piecewise GMM–Class Enumeration

The main modeling task of the unconditional model was to select the optimal growth trajectories and number of classes; recent simulation studies of mixture models have suggested that this unconditional model is the more reliable method for determining the number of classes is to run the class enumeration without the covariates. Class enumeration with covariates (i.e., the conditional model) could lead to poor decisions regarding the number of classes, particularly when the entropy value is lower than 0.80. In some cases, researchers may not want the covariates to influence the determination of the class membership because the inclusion of covariates may potentially change the estimates of class distribution and growth factor means. For determining the number of classes using fit indices, recent simulation studies suggested that BIC performed best among the information criteria and BLRT was proved to be a consistent indicator for deciding on the number of classes (Chen & Kwok, 2009; Nylund-Gibson, 2009; Nylund, et al., 2007).

Following the suggestion of these current developments, the demonstration of the unconditional piecewise GMM herein was geared to optimize the number of classes while choosing a better-fitting growth function. If the fit indices point to inconsistent suggestions on the number of classes, BIC and BLRT will be used as the determinant rules. In addition, the first set of questions A1 through A3 were also addressed at this stage. Note that although the substantive research questions (A1- A3) were posed as distinct and sequential, the actual modeling was executed simultaneously in one single unconditional piecewise GMM. Also note that the variances and covariances structure of the growth factors was specified to the same across classes throughout class enumerations. This is because – when the class-specific variances are allowed – the likelihood function becomes unbound, and because when class-specific covariances between the growth factors are allowed, class separation and interpretation can be comprised.

Question A1: Are there distinct phases where children differ in their speed of learning?

A visual inspection of the observed data displayed in Figures 1 and 2 suggested that a 2-piece linear model summarizes the growth trend better than single linear and quadratic models. Figure 3 displays three latent continuous growth factors: (1) the intercept I representing the starting performance, (2) S1 representing the first growth rate, and (3) S2 representing the second growth rate. Two growth rate factors (i.e., two slopes), in contrast to the traditional one single linear growth rate, were specified to more aptly portray the two-phase growth pattern observed in the data. The two growth rates S1 and S2 depict the non-linear trend by assuming that, within each phase, the growth trajectory was linear. The three growth factors are indicated by part or all of the seven repeated word recognition scores from kindergarten to grade six as shown by the arrows going from the three growth factors in ovals to the seven word recognition scores in rectangles. Note that, at most, a 2-piece linear model was used because each piece requires a minimum of three waves of data, therefore a 3-piece model was not feasible with the study data which has 7 waves; a 3-piece model would require a minimum of 9 waves.

To specify the phasic trajectory, the loadings (i.e., time scores) of the seven measures must be fixed on the three growth factors using the coding scheme often seen in piecewise regression (See Table 2). For the starting performance, the loadings of the intercepts were all fixed at 1. In this demonstration, assuming a grade-2 transition, the loadings of the first growth phase from kindergarten to grade two were fixed at 0, 1, and 2, with an increment of one indicating a constant linear increase across each grade.

In the second phase, the first growth phase loading remained at 2 showing no incremental change to indicate no growth effect in the second phase. The loadings of the second growth phase were fixed at 1, 2, 3 and 4 from grade three to grade six with an increment of one indicating a constant linear increase across each grade. The loadings for S2 were all fixed at zero with no increment indicating no growth effect in the first phase; note that the coding in Table 2 assumed a transition point at the end of grade two. The transition point should be justified by multiple sources of information, including the existing literature, the observed growth trend, the statistical model fit, and the interpretability of the results.

Results for Question A1

First explored was which GMM – single linear, 2-piece linear, or quadratic – fit better by comparing the fit indices. Table 3 shows that the 2-piece models yielded better fit indices than the quadratic models, which in turn fit better than the single linear models irrespective of the number of classes. This indicates that fitting a 2-phase model not only captured the observed non-linearity better than a model merely ignoring the non-linearity but also did better than the commonly used quadratic model. (The default specification for estimation can be found in Chapter 13 of the M*plus* User's Guide.)

The transition point dissecting the two phases was specified at the end of grade two; this decision was made for three reasons. First, the observed pattern shown in Figures 1 and 2 indicated that the transition point occurred at either grade two or grade three. Second, Speece, et al. (2003) studied children from kindergarten to grade three and detected a non-linearity; this suggests that a turning point before grade three was necessary. Also, Francis, et al. (1996) argued that reading difficulty could not be defined at grade one or grade two because identifying reading difficulty often over-

Table 2: Codes for 2-piece Linear Growth Model with Seven Wave of Data

	First Phase			Second Phase			
	K	G1	G2	G3	G4	G5	G6
I	1	1	1	1	1	1	1
S1	0	1	2	2	2	2	2
S2	0	0	0	1	2	3	4

identifies children who have not had adequate educational exposure to reading and under-identifies children who demonstrate deficits in cognitive/linguistic skills. Third, the 2-piece grade-3 transition model yielded negative estimates of growth factor variances, which was undefined and counterintuitive, even with numerous trials of changing the starting values and increasing the number of starting sets up to 1,000 (although the log-likelihood values were replicated). This problem suggested the possibility of an incorrect model. The grade-2 transition models did not show these problems.

Question A2: Are there unknown subpopulations (latent classes) that differ in their growth pattern?

Different learners may display different learning patterns in their reading development. When different groups of learners are empirically observed or theoretically hypothesized, a statistical model must be able to aptly attend to this heterogeneity. Modeling population heterogeneity in growth trajectory is often carried out by a GMM.

GMM is the bedrock of a fully developed GPGMM. GMM relaxes the single population assumption to allow for differences in growth factors across unobserved subpopulations (Kreuter & Muthén, 2007; Muthén, 2004). This flexibility in identifying unobserved subpopulations of people (a.k.a., classes), who are distinct in the developmental pathways, is the cornerstone of the GMM model. The unobserved class membership is modeled by a latent categorical variable where individuals' developmental pathways are relatively similar within each class, yet distinct from one another across classes. As opposed to assuming that individuals vary around a single mean growth trajectory, GMM allows separate mean growth trajectories for each class. Individuals in each class are allowed to vary around the class mean of the growth factors. The variable C in Figure 3 represents such a categorical latent trajectory class.

Results for Question A2

Table 3 compares the results of the 2-piece grade-2 transition models with the number of classes ranging from two to six (rows in bold). The 4-class model was supported by the BIC, the 5-class model was supported by the SBIC, LMR, LRT and BLRT, and the 6-class model was supported by the AIC. With the exception of the 6-class model, all models yielded high entropy values of greater than 0.8. The fit indices point to fairly inconsistent suggestions about the optimal number of classes to extract. The 6-class model was first eliminated from further consideration because it yielded an entropy value lower than 0.80 and because it was suggested only by AIC, which has been shown to be poorer criterion for choosing the correct number of classes (Nylund-Gibson, 2009).

The 4- and 5-class models were each supported by the determinant rule, BIC and BLRT, respectively. To compare the similarities and differences between the 4- and 5-class models, their growth factor means were tabulated and graphed on the first and second panel separately (see Figure 4). Figure 4 shows that the C3 class of the 4-class model branched into two classes of C3a and C3b resulting in 5 classes in total. The 4- and 5-class models were not entirely distinct models; how elaborate the class classification was their main difference. This phenomenon is also common in factor analyses where a model with a greater number of factors is often a more elaborate version of a model with a fewer number of factors.

For demonstrative purposes, it was necessary to select a model with which the conditional piecewise GMM could be demonstrated. In the trial runs of the conditional piecewise GMM, the 5-class model experienced a problem of non-identification and the problem that the log-likelihood could not be replicated – even with the number of starts increased to 10,000 and the assistance of user-specified starting values. For these reasons, the 5-class model was eliminated from further consideration.

The 4-class model, which still revealed a rich substantive story, was chosen because it yielded the smallest value of BIC, which has been shown to be superior in choosing the correct number of classes for GMMs. For real research contexts, choosing the number of classes to extract is not a simple technical task: a researcher must consider multiple factors such as the research purpose, statistical fit and the substantive and practical gains that different numbers of classes may bring about.

Question A3: How are the starting performance and growth rates related?

As discussed in the literature review, there has been a great deal of theoretical and practical interest in whether children with a better start will continue to learn faster and whether children who learn faster at an early age will continue to improve at a faster rate. The field of children's reading development has not settled the debate over how earlier reading performance is related to the later development. In a GMM, these questions are answered by estimating the covariances among the growth factors, I, S1, and S2. These relationships are indicated by the curved arrows among the growth factor I, S1, and S2 in Figure 3.

Table 3: Fit Indices for Single Linear, 2-piece Linear and Quadratic Unconditional Models with 2-6 Classes

		df	L	AIC	BIC	SBIC	LMR LRT	BLRT	Entropy
Single Linear	2-class	15	-10915.215	21856.430	21911.879	21870.613	p=0.169	p=1.000	0.729
	3-class	18	-10996.744	22029.488	22106.399	22049.262	P=0.151	p=0.030	0.489
	4-class	21	-10904.422	21846.844	21927.884	21867.573	P=0.118	p=0.500	0.668
	5-class	25	-10903.928	21851.855	21945.692	21875.858	P=0.365	p=0.800	0.709
2-piece Linear	**2-class**	**20**	**-10015.847**	**20071.695**	**20157.001**	**20093.516**	**p= 0.349**	**p< 0.001**	**0.811**
	3-class	**24**	**-9960.502**	**19969.004**	**20071.372**	**19995.189**	**p< 0.001**	**p< 0.001**	**0.931**
	4-class	**28**	**-9946.927**	**19949.854**	**20069.282**	**19980.403**	**p= 0.065**	**p< 0.001**	**0.880**
	5-class	**32**	**-9936.602**	**19937.204**	**20073.693**	**19972.117**	**p= 0.043**	**p= 0.040**	**0.850**
	6-class	**36**	**-9930.838**	**19933.675**	**20087.226**	**19972.953**	**p= 0.692**	**p= 0.140**	**0.777**
Quadratic	2-class	20	-10227.122	20484.243	20548.223	20500.609	p=0.047	p=1.000	0.682
	3-class	24	-10200.583	20439.166	20520.206	20459.895	p=0.005	p=0.600	0.568
	4-class	28	-10249.835	20543.670	20637.507	20567.673	p=0.099	p<0.001	0.818
	5-class	32	-10211.386	20474.773	20585.671	20503.140	p<0.016	p<0.001	0.823

Notes: df: the number of free parameters of a specified model (when no parameters were fixing to zeros); L: log-likelihood; AIC: Akaike information criterion; BIC: Bayesian information criterion; SBIC: Sample size adjusted BIC; LRT: Lo-Mendell-Rubin adjusted likelihood ratio test; BLRT: bootstrapped parametric likelihood ratio test. The fit indices for the 2-piece models are in bold. The model with the lowest AIC, BIC, or SBIC is underlined. The model with K classes is underlined if the p-value of the LMR LRT or BLRT for the K+1 model was greater than 0.05. When the variance of a growth factor was estimated to be negative, the estimation proceeded with fixing it to zero.

Figure 4: Comparison among the 4- and 5-Class Unconditional and 4-Class Conditional Models

Unconditional 2-piece 4-class Model

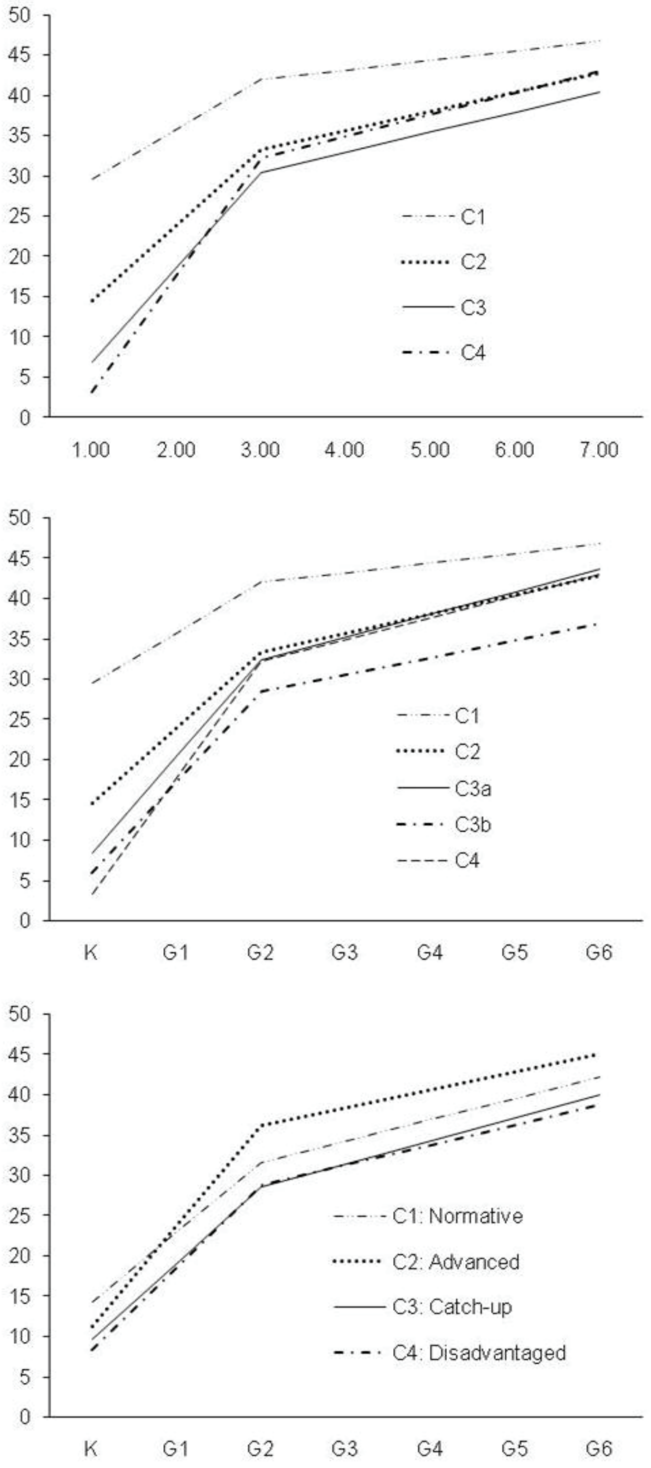
	I	S1	S2	E	%
C1	29.54	6.22	1.20	46.79	0.96
C2	14.53	9.37	2.38	42.79	65.32
C3	6.86	11.80	2.49	40.42	22.82
C4	3.15	14.52	2.72	43.08	10.90

Unconditional 2-piece 5-class Model

	I	S1	S2	E	%
C1	29.56	6.22	1.20	46.79	0.96
C2	14.60	9.35	2.37	42.80	63.81
C3a	8.39	11.98	2.81	43.58	12.10
C3b	5.93	11.24	2.12	36.89	10.67
C4	3.29	14.43	2.72	43.03	12.47

Conditional 2-piece 4-class Model

	I	S1	S2	E	%
C1	14.31	8.69	2.65	42.29	38.33
C2	11.26	12.46	2.21	45.00	33.60
C3	9.72	9.39	2.87	39.97	6.58
C4	8.32	10.24	2.50	38.79	21.49



Notes: I denotes the initial performance at the kindergarten year, S1 denotes the growth rate in the first phase, S2 denotes growth rate in the second phase, E denotes the ending performance at grade six, and % denotes the proportions for the latent classes.

Results for Question A3

By default, M*plus* outputs estimate the covariances of growth factors. For interpretation ease, however, the growth factor correlations were reported by requesting the standardized command in the output. Recall that, for estimation reasons, the covariance structure was fixed to be the same across classes, that is, class-specific correlations among the growth factors were not allowed.

Based on the 4-class unconditional piecewise model, results show that the initial performance was not significantly correlated with the first growth rate ($r = -0.06$, $p = 0.762$), nor was it significantly correlated with the second growth rate ($r = -0.345$, $p = 0.053$). However, the two growth rates were significantly and negatively correlated ($r = -0.599$, $p = 0.001$).

These findings suggest that word recognition performance at the beginning of the kindergarten year, as measured by WRAT-3, was not a good indicator of children's later speed of learning. However, the speed of learning in the first phase may be associated with children's development in the second phase. This suggests that early identification of advanced or disadvantaged children should not rely solely on children's starting performance. Rather, early identification of advanced or disadvantaged children should also look into children's early speed of learning. If a single linear trajectory had been modeled, the relationship between two growth rates would have been overlooked.

Conditional Piecewise GMM with an Auxiliary Developmental Outcome Variable

The conditional piecewise GMM is the full version of GPGMM. It incorporates the covariates and an auxiliary developmental outcome variable into the unconditional piecewise GMM. In this demonstration, five covariates were included. Three were cognitive/linguistic variables that were measured prior to the first assessment of word recognition in the kindergarten year and were standardized scores of verbal working memory, phonological awareness, and word retrieval time. The other two were demographic background variables: gender (boy = 0; girl = 1, 50.2%) and first language reported in the fall of kindergarten year (English = 0, ESL = 1, 15%).

Covariates can have direct and indirect effects on the growth factors. As shown in Figure 3, direct covariate effects explain the growth factor variations, as indicated by the arrow going from the covariates to the growth factors I, S1, and S2. Covariates can also have an indirect effect on the growth factors via their effects on the latent class as indicated by the arrow going from the covariates to the latent class C and then to the growth factors (see Figure 3). The developmental reading outcome variable used in this demonstration was the Stanford Diagnostic Reading Test (SDRT; Karlesen, Madden, & Gardener, 1966) measured at grade six. This developmental outcome variable served as an auxiliary dependent variable for checking the latent class validity, and was standardized for ease of interpretation.

Estimates of class distribution and growth factors means will change as a result of incorporating covariates information and how their effects are specified. Misspecification of the direct and/or indirect effects can lead to untrustworthy estimates. Because the correct population model is unknown and "all models are wrong, the practical question is how wrong they have to be to not be useful" (Box & Draper, 1987, p. 74), it is recommended that researchers experiment with various models. The choice of which model to select must rely heavily on the researchers' discretion borne on the model results (e.g., whether a model terminated normally) as well as their substantive knowledge, and common sense (e.g., checking the tenability of the direction and size of the covariate effects).

In this demonstration, all direct and indirect effects were first allowed on the growth factors for all classes. This model was not identified and the best log-likelihood value was not replicated after numerous trials of starting values and the number of starting values sets = 10,000. Based on the estimated posterior probabilities, this model distributed two very small classes (≈1% and ≈5%; size ≈ 5 and 26), leading to some difficulties in estimating the direct effects on the growth factors (e.g., empty cells in the joint distribution of the model variables). For these reasons, the direct covariate effects of the two small classes were fixed to zeros. This model terminated normally, the log-likelihood values were replicated 3 times with STARTS = 1000 20 (the other 17 differed with their next best values only in the third decimal place). In addition, this model estimated 92 parameters with log-likelihood value = –9693.008. The information criteria of AIC = 19570.016, BIC = 19962.424, SBIC = 19670.392 all (except entropy = 0.768) suggested that this conditional model was a better fitting model than the unconditional model (compare the fit indices of the 2-piece 4-class unconditional model in Table 3).

Note that various models that allowed partial direct effects for the two smaller classes were also examined. Although the log-likelihood values of some of these models were replicated, their the BIC values were all greater than that of the model finally selected and their growth trajectories were harder to recognize and interpret, for these reasons they were not chosen and reported.

The bottom panel of Figure 4 shows the trajectories and estimated growth factor means of the selected conditional model. Two overall observations are pointed out here. First, a noticeable parameter shift was observed when being compared to the unconditional model. The cross-class differences in the growth trajectories diminished a great deal as a result of incorporating the covariates information. Second, it was observed that there was little difference in the estimates of the second growth rates (which only differed in the first decimal place). This result may be a true reflection of the small differences in the speed of recognizing new words among the subpopulations.

The potential ceiling effect of WRAT-3 on the lack of variation in the second growth rate should be considered, however. This instrument lists 42 words for recognition ordered in difficulty and was not originally designed for children. WRAT-3 is known to have a strong ceiling effect (Strauss, Sherman & Spreen, 2006, p. 388). The difficulty level elevates quickly as the words approach the end the list leading to few or no successes in word recognition. This ceiling effect may explain the small class differences in the second growth rates for the present child sample.

The results for specific classes (see the bottom table and graph in Figure 4) indicated that, on average, children in the first class recognized 14.31 words in the kindergarten year, learned 8.69 words per year in the first phase, and 2.65 words per year in the second phase with an ending performance (E) of 42.29 words. This class was referred to the normative class because it consisted of 38.33%, the largest proportion, of the sample, and because its growth trajectory was relatively more typical than those of the other classes.

Children in the second class (33.60%) initially recognized 3.05 fewer words on WRAT-3 than did the normative class. However, their first phase growth rate was 3.77 words faster than the normative class leading to a projection that this class would surpass the normative class at grade one and they would manage to stay ahead of the normative class thereafter despite the slightly slower second growth rate. This class was referred to as the advanced class.

Children in the third class (6.58%) initially recognized 4.59 fewer words on WRAT-3 than did the normative class, but were slowly catching up with the normative class with 0.70 words per year in the first phase and 0.22 words per year in the second phase. This class was referred to as the catch-up class.

The fourth class (21.49%) began with the lowest performance; initially recognizing 5.99 words fewer on WRAT-3 than did the normative class. Although this class appeared to catch up with the normative class with 1.55 words per year in the first phase, they slowed to a rate of 0.15 words per year slower than the normative class in the second phase. In grade six, they recognized 3.50 fewer words than did the normative class and 6.21 fewer words than did the advanced class. This class was referred to as the disadvantaged class.

Question B1: What are the characteristics of the latent classes?

Similar to factors in a factor analysis, the latent trajectory classes do not have inherent meanings (Bauer & Curran, 2003; Kreuter & Muthén, 2007; Muthén, 2003; Muthén, 2004). To understand the characteristics of the latent classes, the categorical latent class variable C is regressed on to the covariates. Covariates play important roles in the GPGMM – they can aid in checking the interpretability and trustworthiness of the selected model. If classes are not statistically different with respect to the covariates, which theoretically or logically should characterize the classes, then there is weak support for the selected model. In Figure 3, the class characteristics regression is shown by the arrow going from the covariates to the latent trajectory class C. Recall that class characterization by covariates can have indirect effects on the growth factors.

Characterization of the latent classes by the set of covariates involves a multinomial logistic regression (or a binary logistic regression for the 2-class case). Coefficients of the covariates in a multinomial logistic regression are linear in the logit form; the increase in the log odds of being in a particular class versus the reference class. The reference class is normally the class with the largest size or the class or that is better recognized by the researchers. The exponent of the slope coefficient, Exp (slope), is the odds ratio for being in one particular class versus the reference class. For example, when comparing ESL (coded as 1) to non-ESL children (coded as 0); a slope = 1 implies that the odds of being in one particular class versus the reference class is Exp (1) = 2.72 times higher for ESL children than non-ESL children.

Results for Question B1

To understand the characteristics of the classes, the results of the multinomial logistic regression were reported and interpreted using the normative class as the reference class. The normative class was used because of its estimated largest proportion and better-known growth pattern. Table 4 reports the slope coefficients (i.e., partial regression coefficient) for the five covariates and their corresponding standard errors and odds ratios. Bear in mind that the interpretation of the odds was based on per one unit change in the covariate. Because the ESL and gender variables were both coded as 0 and 1, their odds reflected the gender and first language differences, and because the cognitive/linguistic variables were all standardized, their odds reflected per SD change.

Relative to the normative class, the odds of membership in the advanced class were significantly increased by being boys. The odds of being in the advanced class versus the normative class were 2.667 (1/odds= 1/0.375) times higher for boys than girls. Relative to the normative class, the odds of membership in the catch-up class were significantly increased by word retrieval time. The odds of being in the catch-up class versus in the normative class were 4.289 times higher per SD increase in word retrieval time. Relative to the normative class, the odds of membership in the disadvantaged class were significantly increased by being a boy, being non-ESL, having poorer phonological awareness and longer retrieval time. The odds of being in the disadvantaged class versus the normative class were 3.106 (1/0.322) times higher for boys than girls, 3.497 (1/0.286) times higher for non-ESL students than ESL students, 2.165 (1/0.462) times higher per SD decrease in phonological awareness, and 1.725 times higher per SD increase in word retrieval time.

Question B2: For each class, what explains children's starting performance and growth rates?

In a GPGMM, the growth factors' variations can also be explained by the same set of covariates. This relationship is analogous to a multiple regression where each of the continuous dependent variables, I, S1 and S2, is regressed onto the covariates. This relationship models the direct effects of the covariates on the growth factors as indicated by the arrows going from the covariates directly to the growth factors. As aforementioned, in the final model only direct effects were allowed on the two classes with larger class proportion, that is, the normative and advanced classes. Note that the indirect effect of covariates on the growth factors via the latent classes, as demonstrated in B2, is reflected by allowing class-varying regression coefficients of the covariates on the growth factors. Thus, the class-varying residual variances were also allowed.

Table 4: Indirect Covariate Effects: Multinomial Logistic Regression for Classes Characterization

	A vs. N	C vs. N	D vs. N
Gender (Girl)	**-0.981** *0.303* 0.375	0.060 *0.810* 1.062	**-1.134** *0.314* 0.322
ESL (Yes)	0.633 *0.389* 1.833	-0.442 *0.842* 0.643	**-1.252** *0.633* 0.286
Verbal Working Memory	0.183 *0.204* 1.201	-0.133 *0.585* 0.875	-0.083 *0.229* 0.920
Phonological Awareness	-0.120 *0.187* 0.887	-0.909 *0.502* 0.403	**-0.773** *0.175* 0.462
Word Retrieval Time	-0.077 *0.245* 0.926	**1.456** *0.374* 4.289	**0.545** *0.231* 1.725

Notes: A: advanced class; N: normative class; D: disadvantaged class; C: catch-up class. Values in the first row of each covariate were the estimates of the slope coefficient, of which the standard errors were placed underneath in italic face, and the corresponding odds ratios were underlined. Significant slope coefficients at the 0.05 level were highlighted in bold face.

Results for Question B2

Results for the class-specific multiple regressions are shown in Table 5. The first row for each covariate reports the estimates of the slope coefficient (partial regression coefficient) and their standard errors were placed underneath in italic face. Significant slope estimates at α= 0.05 level were highlighted in bold. For example, phonological awareness had a significant effect on all growth factors, except for the second growth rate of the normative class. Differential covariate effects in terms of size and direction were found across classes. For example, the initial growth factor I, gender and verbal working memory had significant effects only for the normative class, and word retrieval time had an effect only for the advanced class. Useful substantive information is revealed by comparing differential covariate effects across classes.

Question B3: Do the latent classes differ in the reading developmental outcome?

The GPGMM incorporates an auxiliary outcome variable that can either be proximal or distal. Note that this outcome variable is an auxiliary variable; it is not modeled as an observed dependent variable, nor was it part of the model. Its major role in a GPGMM is to assist in checking the validity of the latent classes by comparing and testing equalities in the class means of this variable (Masyn, 2009; Petras & Masyn, 2009). Because it is an auxiliary variable, the outcome variable is represented in Figure 3 as a dashed square to show that it not an actually modeled outcome variable. This part of the modeling is shown by the arrow going from the latent class variable to the reading comprehension outcome. Cross-class equality in the means of the reading comprehension was tested using the posterior

Table 5: Direct Covariate Effects: Class-specific Multiple Regression of Growth Factors

	Normative			Advanced		
	I	S1	S2	I	S1	S2
Gender (Girl)	**0.565** *0.224*	-0.009 *0.483*	-0.098 *0.197*	0.713 *0.867*	-0.611 *0.493*	-0.011 *0.161*
ESL (Yes)	0.190 *0.272*	**0.971** *0.449*	-0.377 *0.237*	-0.550 *1.125*	-0.813 *0.566*	0.290 *0.219*
Verbal working memory	**-0.258** *0.132*	0.282 *0.226*	-0.107 *0.115*	-0.676 *0.451*	0.148 *0.148*	0.130 *0.094*
Phonological awareness	**0.482** *0.120*	**0.553** *0.234*	-0.147 *0.084*	**4.167** *0.455*	**-1.345** *0.281*	**-0.300** *0.083*
Word retrieval time	-0.240 *0.234*	-0.183 *0.245*	0.146 *0.141*	**-1.714** *0.495*	**0.692** *0.289*	0.119 *0.114*

Notes: Values in the first row of each covariate were the estimates of the slope coefficient, of which the standard errors were placed underneath in italic face. Significant slope coefficients at the 0.05 level were highlighted in bold face.

probability-based multiple imputations method. Since the class means of the reading comprehension were not part of the models, M*plus* needed to estimate means and their asymptotic variances/covariance using the pseudo-class draw technique (Wong, Brown & Bandeen-Roche, 2005). First, individuals' posterior class probabilities (conditional on the model and the observed data) were computed. Next, using this posterior distribution, L pseudo draws were generated for the latent class variable C for individuals – L denotes the number of pseudo draws. Class-specific sample means of the outcome variable then were obtained by taking the average of the L pseudo draws (see M*plus* technical note at http://www.statmodel.com/download/MeanTest1.pdf).

As recommended in Wong, et al. (2005), the M*plus* default number of pseudo draws of 20 was adopted. Equality in the class means were tested using Wald's Chi-square with degree of freedom = K−1 for the omnibus test and 1 degree of freedom for the pairwise tests; a statistically and theoretically/ practically significant and meaningful mean difference should be detected for supporting the validity of the latent class variable. This validity check is analogous to the criterion validity in the traditional psychometrics literature.

Results for Question B3

The last two rows of Table 6 show the class means in the reading comprehension and their corresponding standard errors. First, note that the order of the size of the estimated class means were as expected (i.e., Advanced > Normative > Catch-up > Disadvantaged). The omnibus Wald $\chi^2(3) = 80.094$, $p < 0.001$. The Chi-square values for the paired tests were shown on the upper diagonal of the class matrix in Table 6, and their corresponding *p*-values were shown underneath in italic face. Significant mean differences, highlighted in bold, were found in four of the six paired tests. Mean differences between all non-neighboring classes were all found to be significant (e.g., the difference between the advanced and disadvantaged classes). Mean differences between two of the neighboring classes were found to be non-significant (differences between the advanced and the normative classes and between the catch-up and disadvantaged classes). Judging by the order and size of the class mean estimates and the pattern of the significance tests, the results provided adequate criterion validity evidence for the latent trajectory class variable.

Conclusion

People learn and develop in different ways at different times. These developmental complexities and diversities are often overlooked or modeling tools are incapable of revealing them. This study demonstrated, with children's word recognition development, that by taking into account the phasic learning speed and population heterogeneity, the GPGMM is able to point up evidence for both the deficit and lagging theoretical models reported in literature depending on which classes and developmental phases are being compared.

The advantages of the GPGMM, however, come with a price. To find the optimal model that is both statistically adequate and theoretically interpretable, the GPGMM requires fairly sophisticated modeling techniques that involve iterative and intricate trials of parameter specifications. For a complex model like the GPGMM, the parameter estimates can change remarkably in size and direction from one start set to another. Users are reminded to increase the number of iterations and starting sets when necessary so as to ensure that the log-likelihood of the selected model is replicated. Also, due to the model complexity, the time taken for the estimation to terminate can be much longer than what is needed for simpler models. This is particularly the case when the random start sets are increased to a large number or when the bootstrapped likelihood ratio test is requested. It is suggested that, wherever possible, the GPGMM be run on a spare but fast computer.

To date there is no single agreed-upon best practice for choosing the optimal conditional model. The general statistical problem of choosing the optimal conditional model in latent class models shares a conceptual core in common with indeterminacy problems in factor analysis – note that there are several indeterminacies in factor analysis; for example, indeterminacy of common factors, and an indeterminacy in factor rotation. There may be something to be gained by noting this commonality between latent class and factor analysis. At this point, it is advisable that the unconditional model be used for class enumeration – i.e., for deciding the number of classes. Like the indeterminacy problem of factor rotation, estimates of class distribution and the growth factors of the conditional model may shift from those of the unconditional model

Table 6: Wald's Chi-square Tests of Class Equality in the Means of the Reading Development Outcome

	A	N	C	D
A		0.049 *0.826*	17.688 ***<0.001***	72.170 ***<0.001***
N			16.377 ***<0.001***	60.159 ***<0.001***
C				0.161 *0.688*
M SE	0.273 0.066	0.250 0.076	-0.605 0.196	-0.694 0.092

Notes: A: advanced class; N: normative class; D: disadvantaged class; C: catch-up class. The Chi-square values for the paired tests were shown on the upper diagonal of the class matrix; the corresponding *p*-values were shown underneath in italic face. Significant p-values were highlighted in bold. The class means and their corresponding standard errors were shown in the last two rows.

depending on how the direct and indirect effects are specified. Recent work by Nylund-Gibson (2009) suggests that first the indirect effects be added to the unconditional model followed by the direct effects. The parameter shift then is examined throughout the steps. Implicitly, this suggestion was used along with verifying the substantive interpretability, as a rough guide for checking and selecting a conditional model.

An intuitive, yet less than ideal solution, is to fix the growth factor parameters of the conditional model to those estimated by the unconditional model. By doing so, the covariate effects can be investigated without shifting the growth factor parameter estimates. This method could be problematic because it treats the fixed parameters as if they were true population values and ignore the sampling errors. Moreover, using the estimates of the unconditional model for the conditional model may be considered as cheating because both models are based on the same sample set. Hence, this strategy is not recommended if the purpose of the GPGMM is of an exploratory nature as demonstrated in this study. It may be more justified if the purpose is to cross-validate, that is, to verify growth trajectories suggested by other samples from the same population.

Traditionally, questions B1, B2, and B3, as addressed by the conditional model, are often answered by saving the likely class membership or the posterior probabilities for each individual in a new data file and running separate analyses. This method could also be problematic because the class membership or the posterior probabilities are treated as being observed, but they are, in fact, model estimates with errors. Recent studies have shown that these traditional approaches may yield incorrect parameter estimates and standard errors leading to incorrect conclusions about significance testing (Clark & Muthén, 2009; Masyn, 2009; Petras & Masyn, 2009). Answering these questions using a single GPGMM circumvents this problem, especially when the entropy is high (Clark & Muthén, 2009).

A trade-off between the number of classes extracted and the amount of variance of the growth factors (or residual variance after adding the covariates) was noticed. This phenomenon makes sense conceptually and statistically because the mechanism behind the GMM is to extract K classes where people are relatively similar within each class, yet distinct from one another across classes.

In a highly hypothetical situation where K is equal to the sample size, there will be no within-class variation in the growth factors. The 4- and 5- class conditional models encountered scenarios where the variances and/or residual variances of the growth factors being estimated were negative and received warning messages such as non–positive definite latent variable covariance matrix. Fixing the negative residual variances to zero may solve these problems, however, these problems may be indicative of class over-extraction or misspecification of the covariate effects – this is conceptually similar to a Heywood case in factor analysis.

The balance between number of classes and the within-class variances/residual variances often dictates the number of classes one is able to interpret, especially for the conditional model. The maximum number of interpretable classes is often bounded by how much variance the growth factors are estimated to have and whether the variance is sufficient for the conditional model. Using the study data, difficulty in identifying the 5-class conditional model was experienced, although it is preferred for more richness in the substantive information.

With a full GPGMM, a large number of parameters are simultaneously estimated and the number of parameter estimates increases rapidly in multiples as the number of classes and covariates increase. The large set of the parameters is deemed to be the best solution for the data simply because it yields the least -2 log-likelihood value. The maximum likelihood algorithm cannot tell whether or not the parameter estimates, in term of size and direction, make sense for a real and specific research context. Valid interpretations of the GPGMM results rely heavily on the users' methodological and substantive knowledge of the study. This demonstration showed that the speed of learning new words slowed down in the second phase for all classes; however, it would be inappropriate to conclude that children learn fewer words annually after grade two than before grade two without some special caution. As mentioned, this finding may result from the low floor effect but strong ceiling effect of the WRAT-3. As stated by Muthén (2003) and stressed throughout this article, a quality GPGMM should be guided not only by the statistical information, but also by the substantive interpretability of the results. GPGMM, in essence, is merely an analytical tool. Substantive expertise throughout the process of model specification, selection, and verification is the key to the success of a GPGMM.

References

Aarnoutse, C., van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing*, *14*, 61-89.

Aunola, K., Leskinen, E., Onatsu-arvilommi, T., & Nurmi, J-E. (2002). Three methods for studying developmental change: A case of reading skills and self-concept. *British Journal of Educational Psychology*, *72*, 343-64.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual difference in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, *34,* 1373-1399.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumption of growth mixture models: Implications for over extraction of latent trajectory classes. *Psychological Methods, 8*, 338-63.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology*, *100*, 192-208.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York: Wiley.

Chen, Q., & Kwok, O. (2009). Can the correct number of classes be detected in misspecified multilevel growth mixture models? A Monte Carlo study. Paper presented at the 2009 Annual Meeting of American Educational Research Association (AERA), San Diego, CA.

Clark, S., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. *Submitted for publication*.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., & Shaywitz, B. (1996). Development lag versus deficit models for reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology*, *88*, 3-17.

Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Finding of two-year longitudinal study. *Journal of Educational Psychology*, *94*, 586-97.

Karlesen, B., Madden, R. and Gardener, E.F. (1966). *Stanford diagnostic reading test*. New York: Harcourt, Brace & World.

Khoo, S. T. (1997). *Assessing interaction between program effects and baseline: A latent curve approach*. Unpublished doctoral dissertation, University of California, Los Angles.

Kreuter, F. & Muthén, B. (2007). Longitudinal modeling of population heterogeneity: Methodological challenges to the analysis of empirically derived criminal trajectory profiles. In *Advances in latent variable mixture models*, G. R. Hancock, & K. M. Samuelsen (Eds.). Charlotte, NC: Information Age Publishing.

Leppänen, U., Niemi, P., Aunola, K., & Nurmi, J-E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly*, *39*, 72-93.

Li, F., Duncan T. E., Duncan, S. C., & Hops, H. (2001). Piecewise growth mixture modeling of adolescent alcohol use data. *Structural Equation Modeling*, *8*, 175-204.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767-78.

Masyn, K. E. (2009). *The consequences of (latent) classes: Results of a simulation study evaluating approaches to specifying and testing the effect of the latent class membership on distal outcomes*. Paper presented at the 2009 Annual Meeting of American Educational Research Association (AERA), San Diego, CA.

McGee, V. E., & Carleton, W. T. (1970). Piecewise regression. *Journal of the American Statistical Association*, *65*, 1109-1124.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Meredith, W., & Tisak, J. (1990). Latent curve analysis, *Psychometrika*, *55*, 107-122.

Muthén, B.O. (1998-2004). *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide* (*5th Ed.*). Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using EM algorithm. *Biometrics*, *55*, 463-9.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*, 882–891.

Muthén, B. (2001). Second generation structural equation modeling with a combination of categorical and continuous latent variable: New opportunities for latent class/latent growth modeling. In *New methods for analysis of change*, L. M. Collins & Sayer (Eds.), 291-232). Washington, DC: APA.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 369-77.

Muthén, B., Khoo, S. T., Francis, D., & Kim Boscardin, C. (2003). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In *Multilevel modeling: Methodological advances, issues and applications*, S.R. Reise & N. Duan (Eds.), 71-89. Mahaw, NJ: Lawrence Erlbaum Associates.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In *Handbook of quantitative methodology for the social sciences*, D. Kaplan (Ed.), 345-368. Newbury Park, CA: Sage Publications.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In *Advances in latent variable mixture models*, G.R. Hancock & K. M. Samuelsen (Eds.), 1-24. Charlotte, NC: Information Age Publishing.

Muthén, B., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In *Longitudinal data analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), 143-165. Boca Raton, FL: Chapman & Hall/CRC Press.

Nylund, K. L., Asparouhov, T., Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(*4*), 535-569.

Nylund-Gibson, K. L. (2009). *Covariates and mixture modeling: Results of a simulation studying exploring the impact of misspecified covariate effects*. Paper presented at the 2009 Annual Meeting of American Educational Research Association (AERA), San Diego, CA.

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J-E, & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finish. *Journal of Educational Psychology*, *97,* 299-319.

Petras, H., & Masyn, K. (2009). General growth mixture analysis with antecedents and consequences of change. In *Handbook of quantitative criminology*, A. Piquero & D. Weisburd (Eds.), 69-100. New York: Springer.

Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through six grades. *Journal of Educational Psychology*, *94*, 3-13.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (*2nd Ed.*). Newbury Park, CA: Sage.

Scarborough, H. S., & Parker, J. D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psychological problems form Grade 2 to Grade 8. *Annals of Dyslexia*, *53*, 47-71.

Speece, D. L., Ritchey, K. D., Cooper, D. H., Roth, F. P., & Schatschneider, C. (2003). Growth in early reading skills form kindergarten to third grade. *Contemporary Educational Psychology*, *29*, 312-332.

Strauss, E., Sherman, E., & Spreen, O. (2006). *A compendium of neuropsychological tests: administration, norms and commentary* (*3rd Ed.*). New York: Oxford University Press.

Wilkinson, G. S. (1995). *The wide range achievement test-3*. Wilmington, DE: Jastak Associates.

Wong, C-P, Brown, C. H., & Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association*, *100*, 1054-1076.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson psychoeducational test battery*. Boston: Teaching Resources.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis*, *50*, 1090-1104.

Appendix:
M*plus* Syntax for the Final GPGMM Conditional Model

```
TITLE: GPGMM WORD RECOGNITION

DATA:
   FILE IS wrat526.dat;
   FORMAT IS 418F22.0;

VARIABLE:
   NAMES ARE K G1 G2 G3 G4 G5 G6
   Gender FirstLanguage
   WorkingMemory PhonoAwareness
   RetrievalTime ReadingComprehension;
   CLASSES = C(4);
   MISSING = K G1 G2 G3 G4 G5 G6
   ReadingComprehension (9999);
   AUXILIARY= (e)ReadingComprehension;
   USEVAR = K G1 G2 G3 G4 G5 G6 Gender
   FirstLanguage
   WorkingMemory PhonoAwareness
   RetrievalTime ReadingComprehension;

ANALYSIS:
   TYPE = MIXTURE;
   STARTS = 1000 20;
   STITERATIONS = 100;
   MITERATIONS = 2000;

MODEL:
   %OVERALL%
   I S1 | K@0 G1@1 G2@2 G3@2 G4@2
   G5@2 G6@2 ;
   I S2 | K@0 G1@0 G2@0 G3@1 G4@2
   G5@3 G6@4 ;
   C#1 C#2 C#3 on Gender*0 FirstLanguage*0
   WorkingMemory*0 PhonoAwareness*0
   RetrievalTime*0;
   I ON Gender*0 FirstLanguage*0
   WorkingMemory*0 PhonoAwareness*0
   RetrievalTime*0;
   S1 ON Gender*0 FirstLanguage*0
   WorkingMemory*0 PhonoAwareness*0
   RetrievalTime*0;
   S2 ON Gender*0 FirstLanguage*0
   WorkingMemory*0 PhonoAwareness*0
   RetrievalTime*0;

   %C#1%
   K@0 G1 G2 G3 G4 G5 G6
   I S1 S2@0;
   [I*14.326 S1*8.665 S2*2.612;
   I with S1 @0;
   I        ON
   Gender*0.597
   FirstLanguage*0.175
   WorkingMemory*-0.221
   PhonoAwareness *0.470
   RetrievalTime*-0.185;
   S1       ON
   Gender* -0.007
   FirstLanguage* 0.907
   WorkingMemory*0.300
   PhonoAwareness *0.500
   RetrievalTime*-0.162;
   S2       ON
   Gender*-0.074
   FirstLanguage*-0.375
   WorkingMemory*-0.081
   PhonoAwareness*-0.141
   RetrievalTime*0.183;

   %C#2%
   K@0 G1 G2 G3@0 G4 G5 G6 ;
```

```
I S1 S2@0;
[I*8.436 S1*10.289 S2*2.911];
I with S1 @0;
I        ON
Gender@0
FirstLanguage@0
WorkingMemory@0
PhonoAwareness@0
RetrievalTime@0;
S1       ON
Gender@0
FirstLanguage@0
WorkingMemory@0
PhonoAwareness@0
RetrievalTime@0;
S2       ON
Gender@0
FirstLanguage@0
WorkingMemory@0
PhonoAwareness@0
RetrievalTime@0;

%C#3%
K* G1 G2 G3 G4 G5 G6;
I S1 S2@0;
[I*11.329 S1*11.352 S2*2.156];
I with S1@0;

I        ON
Gender*0.338
FirstLanguage* 0.177
WorkingMemory*-0.745
PhonoAwareness*4.112
RetrievalTime*-1.645;
S1       ON
Gender*-0.524
FirstLanguage*-0.914
WorkingMemory*0.146
PhonoAwareness*-1.250
RetrievalTime*0.659;
S2       ON
Gender*0.025
FirstLanguage*0.215
WorkingMemory*0.141
PhonoAwareness*-0.304
RetrievalTime*0.055;
```

```
     %C#4%
     K G1 G2 G3 G4 G5 G6@0;
     I S1@0 S2;
     [I*8.386 S1*10.041 S2*2.513];
    I WITH S2@0;
    I        ON
    Gender@0
    FirstLanguage@0
    WorkingMemory @00
    PhonoAwareness@0
    RetrievalTime@0;
    S1       ON
    Gender@0
    FirstLanguage@0
    WorkingMemory@0
    PhonoAwareness@0
    RetrievalTime@0;
    S2       ON
    Gender@0
    FirstLanguage@0
    WorkingMemory @0
    PhonoAwareness@0
    RetrievalTime@0;

OUTPUT:
     TECH1 TECH4;
     !STANDARDIZED;
     SAVEDATA:File is final.sav;
     SAVE = FSCORES;
```