Use of Hierarchical Linear Modeling and Curriculum-Based Measurement for Assessing Academic Growth and Instructional Factors for Students with Learning Difficulties

Jongho Shin
Seoul National University
Korea

Christine A. Espin
University of Minnesota
U.S.A

Stanley L. Deno
University of Minnesota
U.S.A

Scott McConnell
University of Minnesota
U.S.A

The main purpose of this paper is to demonstrate how to apply the Hierarchical Linear Modeling (HLM) technique to multi-wave Curriculum-Based Measurement (CBM) measures in modeling academic growth and assessing its relations to student- and instruction-related variables. HLM has advantages over other statistical methods (e.g., repeated measures ANOVA, Structural Equation Modeling) in modeling academic growth. The advantages include allowing more flexible research designs in collecting multiple data points and estimating growth rates and their relations to correlates in more reliable, accurate ways. CBM, as a multi-wave progress-monitoring system, also has distinctive psychometric features that facilitate longitudinal research on academic skill development. These features include provision of multiple data points within short time periods, good validity and reliability, and sensitivity for detecting small degrees of change. Finally, research questions related to assessing the academic growth of students with learning difficulties and using assessment results to improve educational practices for them are discussed

Key Words: growth modeling, Hierarchical Linear Modeling (HLM), Curriculum Based Measurement (CBM)

A major goal of education has been to produce changes in the knowledge and skill levels of students. Although measuring change over time is important for all educators, it is especially important for teachers who are in charge of students with learning difficulties and must monitor individual students' progress over time and evaluate the effects of instructional programs for those students. The academic growth of students with learning difficulties, unfortunately, is often so minimal as to go undetected by typical published, standardized achievement tests, especially over short periods of time. In addition, it is not easy to find

testing instruments that generate reliable and valid repeated measures of student performance over time, especially within a short time period (e.g., one year). For example, published standardized achievement tests are designed to be given at approximately one-year time intervals; they are not designed for multiple administrations within a one-year time frame.

Over the past 20 years, a system of measurement referred to as Curriculum-Based Measurement (CBM) has been developed that can be used to measure student performance reliably, validly, and repeatedly to represent academic growth over time (Deno, 1985; Good & Jefferson, 1998; Marston, 1989; Shin, Deno, & Espin, 2000). CBM has been used primarily as a method for teachers to evaluate the progress of individual students over time so that teachers can evaluate the effectiveness of their instruction. However, CBM also has great potential for enabling researchers to examine the academic growth patterns and rates for groups of students, and further, to examine the relationship between student growth and relevant correlates (i.e., variables

Jongho Shin, Assistant Professor at the Department of Education of the Seoul National University.
Christine Espin, Stanley Deno, and **Scott McConnell**, Professors at the Department of Educational Psychology of the University of Minnesota. Correspondence concerning this article should be addressed to Jongho Shin, Department of Education, Seoul National University, San 56-1 Shinrim-Dong Kwanak-Gu, Seoul 151-748, Korea. Electronic mail may be sent to jshin21@snu.ac.kr.

assumed to be associated with student growth such as homework, class participation, and motivation to learn).

Repeated measurement of student performance, a hallmark characteristic of CBM, allows educators to make instructional decisions based on student growth over time rather than student status at a given time. CBM, as a multi-wave growth monitoring system, has technical, psychometric advantages in assessing student growth and its relations to correlates. In that regard, CBM holds promise for assessing the slow development of students with learning difficulties (Deno, 1985; Deno & Fuchs, 1987; Marston, Deno, & Tindal, 1983).

The main purpose of this paper is to demonstrate how to use CBM as a multi-wave growth monitoring system for assessing student growth in combination with a statistical method called Hierarchical Linear Modeling (HLM) (Bryk & Raudenbush, 1987, 1992). Before describing specific procedures for modeling student growth using CBM and HLM, we briefly present (a) psychometric features of CBM for modeling academic growth and (b) statistical methods that handle multiple data points for assessing student growth over time (i.e., HLM, Structural Equation Modeling, and repeated measures ANOVA). We then demonstrate how to model student growth and how to examine the effects of student- and instruction-related variables on student growth using multi-wave CBM math measures with HLM. Finally, we discuss research questions related to the investigation of academic growth of students with learning difficulties using CBM procedures.

CBM and Its' Psychometric Features

CBM has been used to monitor students' progress in the academic skills of reading, mathematics, and writing, and its' technical adequacy in validity and reliability has been well established (Deno, 1985; Marston & Magnusson, 1985). Students' academic performance is assessed with testing items developed on the basis of school curricula. Equivalent forms of tests in each skill area are provided for measuring students' progress over time, and their testing scores are graphed and provided for teachers and students themselves.

The first advantage of using CBM for assessing student growth relates to the technical adequacy of the measures (Good & Jefferson, 1998; Marston, 1989). The validity and reliability of performance measures are important if tests are to be used in educational decision-making, as well as in the study of growth. An extensive body of research shows that

CBM measures as static measures (e.g., measures obtained on a given testing occasion) have strong criterion-related validity with standardized achievement tests (Good & Jefferson, 1998; Marston, 1989). Further, the *growth rates* estimated on repeated CBM measures collected over time have good predictive validity on later student achievement on standardized tests (Shin, Deno, & Espin, 2000).

With regard to reliability, research shows that most reliability coefficients of CBM measures are above .90 (Marston, 1989). This high reliability indicates that CBM measures can be used to make educational decisions on individual students for identification purposes (Salvia & Ysseldyke, 1995). Related to the study of student growth, the high reliability of CBM measures increases the reliability of estimated growth parameters by decreasing the measurement error. The increased reliability results in more dependable examination of the relations between growth parameters and correlates (Willet, 1989b).

The second advantage of using CBM measures for the study of student growth is its logistical efficiency in obtaining multiple data points over short time periods. As the number of data points increases, the measurement error associated with growth estimates decreases, leading to higher reliability of growth parameters (Willet, 1989a, 1989b). Subsequently, increased reliability of growth parameters enables more accurate examination of the effects of students' background variables (both static, like SES, and variable, like motivation) on academic growth.

In addition, various growth trajectories (e.g., a quadratic growth model that can estimate acceleration or deceleration of growth rates over time) can be examined by using multiple data points. In contrast, with two data points, student growth must be assumed to be linear (Willet, 1989a); however, psychological or behavioral traits of human beings rarely develop at a constant rate over time.

The third advantage of using CBM for assessing student growth is its sensitivity to changes over short periods of time. Using sensitive measures is important, especially when timely instructional decisions for individual students are to be made on the basis of student growth. Research shows that CBM measures are more sensitive to changes in student performance over short time periods than standardized achievement tests (Marston, Deno, & Tindal, 1983; Marston & Magnusson, 1985), and to inter-individual differences in rates of growth over time (Shin, Deno, & Espin, 2000).

Finally, CBM employs behavioral measures of student performance (e.g., number of words read correctly per minute) such that individual differences in academic

performance are allowed to increase over time. Individual differences in academic skill development are likely to increase over time because every individual does not develop academic skills at the same rate (Labouvie, 1982); therefore, a measure for assessing student growth should be sensitive to such increase of individual differences over time. Increased variances in test scores across time lead to increased variability in individual growth rates across time. Subsequently, such increased heterogeneity of growth rates among individuals is expected to result in higher reliability of growth-rate estimates (Bryk & Raudenbush, 1987, 1992; Willet, 1989b; Zimmerman & Williams, 1982).

Statistical Methods Handling Multiple Data Points for the Study of Student Growth

Using multi-wave data points in the study of academic growth provides (a) more reliable estimation of growth parameters, (b) a method for examining growth patterns, and (c) a method for more systematically investigating relations between growth parameters and correlates. Statistical methods that handle multiple data points for the study of student growth include repeated measures ANOVA (Hertzog & Rovine, 1985; McCall & Appelbaum, 1973), Structural Equation Modeling (SEM) (Loehlin, 1998; Maruyama, 1998; Willet & Sayer, 1994), and Hierarchical Linear Modeling (HLM) (Bryk & Raudenbush, 1987, 1992). Each of these statistical methods has distinctive technical characteristics related to statistical assumptions, intervals between testing occasions, missing-data handling, characteristics of correlates, and the number of subjects required to get reliable growth estimates. In this section, we describe and compare each of the three statistical methods for the study of student growth based on multi-wave performance data.

Repeated Measures ANOVA

Repeated Measures ANOVA is a statistical method that deals with differences among sample means obtained from the same participants. Using this method, researchers can examine developmental patterns of student performance over time and the effect of correlates on performance changes based on repeated measures of student performance. In this case, repeated measures become a within-individual variable and correlates (e. g., high-, average-, and low-achieving students) become a between-individual variable.

Repeated Measures ANOVA requires meeting strict assumptions on repeated measures of student performance.

First, the method requires relations among repeated measures of student performance to be consistent across testing occasions (i. e., compatible variance and covariance matrixes), called the "sphericity" assumption (Hertzog & Rovine, 1985; McCall & Appelbaum, 1973). In reality, the relation between closely located measures is likely to be different from the relation between distantly located measures. When the sphericity assumption is not met, either correction procedures (i.e., adjusting the degree of freedom for statistical tests) or multivariate statistics (e. g., Wilk's λ) is suggested as an alternative to ordinary univariate F tests (Hertzog & Rovine, 1985).

A second assumption of Repeated Measures ANOVA is that repeated measures of student performance are obtained at the same time for all participants with equal intervals between testing occasions. Repeated Measures ANOVA, therefore, is not appropriate for repeated measures obtained at different time points among students or with different time intervals between testing. When conducting frequent repeated measurement of students over time (e. g., over an entire school year), it becomes difficult to test all students on the same occasions and with the same time intervals between testing. Therefore, the flexibility of research design for the study of student growth is decreased when using Repeated Measures ANOVA.

An additional problem with Repeated Measures ANOVA is the way in which missing data are handled. If cases have any missing data, they are automatically excluded from analysis. In reality, when students are repeatedly assessed over time (e. g., an academic year), it is likely some students will miss a test on a certain occasion. Eliminating all students who have incomplete data reduces the statistical power to detect relations between student growth and correlates, resulting in misrepresentative outcomes.

Finally, only discrete variables (e. g., group classification, gender, and ethnicity) can be used in Repeated Measures ANOVA as correlates (predictors) to explain inter-individual differences in performance changes. Therefore, when continuous variables (e. g., motivation, active responding to teacher's requests, and age) are considered as a correlate, other statistical methods (e. g., SEM and HLM) must be used.

Structural Equation Modeling

Structural Equation Modeling (SEM) is a regression-based statistical method that deals with relationships among measures including repeated measures of student performance. Similar to Repeated Measures ANOVA, SEM

requires strictly defined repeated performance data (Loehlin, 1998; Maruyama, 1998; Willet & Sayer, 1994). First, performance data must be collected with equal time intervals between testing occasions over time. If testing occasions are not evenly distributed, the reliability and accuracy of growth estimation is reduced. Second, as with Repeated Measures ANOVA, SEM requires all individuals to have complete data. That is, cases having missing data are automatically eliminated from analysis.

In contrast to Repeated Measures ANOVA, SEM enables researchers to use both discrete and continuous variables as correlates to examine factors associated with higher growth rates over time. In addition, multiple indicators can be used to estimate *true* (more reliable) performance levels at each time point (Loehlin, 1998; Maruyama, 1998). CBM easily produces multiple testing scores in a testing occasion; therefore, these multiple indicators can be used in SEM to estimate more stable student performance, resulting in more accurate and dependable growth rates.

Large sample sizes, however, are required when SEM is used for assessing student growth and examining its relations to correlates. When small sample sizes are used, estimation of growth parameters and examination of relations between student growth and correlates become less reliable (Loehlin, 1998).

Hierarchical Linear Modeling

Hierarchical Linear Modeling (HLM) is a regression-based statistical method that deals with multi-level data including repeated measures of student performance (i.e., repeated scores nested within individual students) (Bryk & Raudenbush, 1992). HLM enables researchers to examine students' academic skill development using more flexible and practically plausible research designs than those possible with Repeated Measures ANOVA or SEM. First, HLM does not require all individual students to be tested at the same time points. Student performance data can be collected on different time schedules, which is often necessary for large-scale assessment at a school, district, or state level. Second, HLM efficiently handles missing data. Each testing occasion for an individual student is treated as a separate case so that only missing data points, not individuals having missing data, are excluded from the analysis. Third, researchers can use both continuous and categorical predictors when employing HLM to examine relations between growth rates and correlates. Furthermore, HLM allows differential weights to be used in examining relations between growth rates and

correlates (Bryk & Raudenbush, 1987, 1992). For example, individuals whose growth rates are estimated more reliably (i.e., smaller standard errors of estimation of growth rates) are given higher weights than those with less reliable estimates. Finally, HLM allows for a relatively small number of students to be used to estimate growth parameters, in contrast to SEM (Bryk & Raudenbush, 1992). For these reasons, HLM appears to be a better tool for examining academic skill development and its relations to correlates than other statistical methods.

Application of Hierarchical Linear Modeling to the Study of Growth Monitoring

In this section, we demonstrate how educators and educational researchers can use HLM to examine patterns and rates of students' academic skill development and to identify instructional factors facilitating student growth over time. In HLM, the examination of academic growth based on multiple data points is conceptually divided into two different stages: within- and between-individual stages (Bryk & Raudenbush, 1987, 1992; Raudenbush & Bryk, 1989). The main focus of the within-individual stage is on (a) identifying an appropriate growth trajectory and then (b) estimating growth parameters on the basis of a selected growth trajectory. This contrasts with the between-individual stage where emphasis shifts to testing models that account for individual differences in growth rates with correlates as predictors.

Description of Data Used for Demonstration

For the purposes of demonstration, CBM math data were used that had been collected monthly from November 1998 to May 1999 in an elementary school in the Midwest of the United States. A main purpose of the data collection was to examine the effects of the computer-based, instructional system called Discourse (see Shin, Deno, Robinson, & Marston, 2000, for detailed description) on students' participation in class activities and academic achievements.

The CBM math data were from 63 third graders and 74 fourth graders, whose basic skills of addition, subtraction, multiplication, and division were tested through the Discourse system. One minute was allowed to students for the test, and the number of problems answered correctly was used for analysis. Means and standard deviations of CBM math scores on each testing occasion were reported in Table 1.

Fifteen participants (10.9%) were those having learning

Table 1. *Means and Standard Deviations of Monthly CBM Math Scores*

	November	December	January	February	March	April	May
Grade 3	7.77 (5.90)	10.04 (6.45)	13.04 (6.26)	13.72 (7.21)	14.54 (7.55)	16.17 (7.73)	18.80 (10.39)
Grade 4	14.95 (7.30)	15.23 (7.87)	19.90 (7.57)	22.91 (9.66)	24.47 (11.50)	30.25 (16.75)	29.16 (15.02)
All	11.54 (7.56)	12.90 (7.69)	16.74 (7.77)	18.65 (9.73)	20.06 (11.08)	23.82 (15.09)	24.03 (13.89)

Note. Numbers in parenthesis are standard deviations.

Table 2. *Means and Standard Deviations of Scaled Scores (SS) and Normal Curve Equivalents (NCE) of Math, Reading, and Complete Batteries of the Metropolitan Achievement Tests-7*

	Math battery		Reading battery		Complete battery	
	SS	NCE	SS	NCE	SS	NCE
Grade 3	540.93 (37.46)	44.92 (23.29)	565.19 (48.27)	49.56 (22.92)	558.74 (34.46)	46.04 (23.29)
Grade 4	568.16 (35.72)	46.08 (21.24)	575.87 (41.72)	40.08 (20.75)	576.03 (33.25)	41.50 (21.03)
All	555.84 (38.83)	45.55 (22.11)	571.04 (44.93)	44.37 (22.18)	568.21 (34.76)	43.56 (22.11)

Note. Numbers in parenthesis are standard deviations.

disabilities (LD), and 70 students (62.5%) received free or reduced lunch in the school. Forty-eight percent of the participants were female and 52% were male. They took the Metropolitan Achievement Tests-7 (MAT-7) at the beginning of the 1998/99 school year. Descriptive statistics of the MAT-7 reading, mathematics, and complete battery scores were displayed in Table 2.

Within-Individual Stage: Growth Pattern Examination

During the within-individual stage, an appropriate growth model is usually selected by using theoretical hypotheses of developmental patterns, visually inspecting individual growth curves, and examining the goodness of fit to existing data (Willet, 1989a). Growth models in which growth parameters are interpretable, however, should be selected. For example, less than a cubic term would be recommended because linear and quadratic terms can be interpreted meaningfully as constant rates of change and positive or negative acceleration of linear rates, respectively (Shaywitz & Shaywitz, 1994).

Using the CBM math data for students with and without learning disabilities (LD), we considered a developmental lag hypothesis as a theoretical rationale for selecting a growth model (Stanovich, Nathan, & Zolman, 1988). According to this hypothesis, students with LD develop math skills slowly at the beginning stage of learning and then develop the skills at faster rates than their peers without disabilities do. This

hypothesis provides a theoretical basis for selecting a quadratic growth model, as follows:

$$Y_{it} = \pi_{0i} + \pi_{1i} \times a_{it} + \pi_{2i} \times a_{it}^2 + r_{it},$$

where Y_{it} is an observed score at time t for individual i, π_{0i} the intercept, π_{1i} the linear growth rate at the time point of zero (i. e., intercept), π_{2i} the quadratic growth rate indicating acceleration of the linear growth rate, a_{it} time of data collection, and r_{it} the random error (Bryk & Raudenbush, 1992).

Visual inspection of individual students' growth curves provides another basis for determining an appropriate growth model. Individual students' growth curves displayed in Figure 1 suggest that students' math scores increase over time, but that the amounts of change appear to decrease slightly toward the end of the school year.

A statistical test can be used to examine the degree of model fit to data in HLM, called the likelihood-ratio test on deviance statistics (Bryk & Raudenbush, 1992). HLM reports a deviance statistic of a growth model selected to estimate growth parameters. The deviance statistic indexes the degree of model fit to data. The higher the deviance statistic, the poorer the model fit. The difference between deviance statistics of different growth models (e. g., linear and quadratic models), therefore, serves as an indicator of how much we could increase the degree of model fit to data by choosing a certain growth model. Furthermore, the difference

of deviance statistics between different growth models can be statistically tested because the difference score is known to have a χ^2 distribution with the degrees of freedom equal to the difference in the number of estimated parameters between two models (Bryk & Raudenbush, 1992).

For the purposes of demonstration, deviance statistics were examined for linear and quadratic growth curves delineating students' math-skills development. The deviance statistic of a linear growth model was 4756.49, whereas the value for the quadratic model was 4717.44. A smaller deviance statistic for the quadratic model suggests that it might be a better fit to the data than a linear model. Furthermore, a statistical test using χ^2 statistics indicates that the difference of deviance statistics (i. e., difference score of 39.05) between the two models was statistically significant ($\chi^2 = 39.05$, $df = 3$, $p < .01$). Based on the result, the quadratic model was adopted for further analyses.

Within-Individual Stage: Growth Rate Estimation

After selecting an appropriate growth model, research interest shifts to estimating group means and variances of growth parameters (e.g., initial status, linear growth rates, and acceleration of linear rates in a quadratic growth model), as well as individual students' growth parameters. In HLM, group-mean estimates of growth parameters are referred to as "fixed effects," whereas variance estimates (e. g., variations of intercepts and growth rates among students) are referred to as "random effects." HLM reports estimates of fixed and random effects (i. e., group means and variances) when a selected growth model is applied (see Table 3).

Table 3 shows that mean performance toward the beginning of the school year (i. e., November of 1998) for all students was 11.58 problems answered correctly, which was

statistically different from the initial status of zero. In addition, students as a group showed a positive linear growth rate of 2.62 problems increased per month. The linear growth rate estimated was also statistically different from the growth rate of zero. In contrast, the group mean of quadratic growth rates was statistically non-significant, though the group-mean linear growth rate looked to decrease at a rate of .06 problems per month. The result of the statistical non-significance did not mean that the quadratic growth model would not fit the data. Even though the group mean of quadratic growth rates was not significant, there would be significant inter-individual differences in quadratic growth rates. In this case, the quadratic model should be adopted.

Examination of random effects (see bottom of Table 3) shows that significant inter-individual differences (i. e., variance) were found in initial status, linear growth rates, and quadratic growth rates. In other words, these variances tell us that individual differences existed in the levels of math basic skills at the start of the school year (i. e., initial status), in the simple rates of growth, and in the amounts of deceleration of linear growth rates among individual students.

The "level 1 error" of random effects (see bottom of Table 3) indicates the average prediction error of individual students' observed scores using a quadratic growth model. This index of average prediction error also can be used to determine which growth model better depicts developmental patterns of individual students. For example, in the CBM math data used for demonstration, the average prediction error was smaller for the quadratic model (i. e., 23.32) than that for the linear model (i. e., 27.15). Again, the results suggest that a quadratic model better fit the CBM math data.

Between-Individual Stage: Test of Developmental Lag Hypothesis

Table 3. *Group Mean and Variance Estimates of Growth Parameters in a Quadratic Growth Model*

Fixed effect	Coefficient	SE	t ratio	p value
Initial status (β_{00})	11.58	0.68	17.08	.00
Linear slope (β_{10})	2.62	0.39	6.69	.00
Quadratic slope (β_{20})	-.06	0.07	0.81	.42
Random effect	Variance component	df	χ^2	p value
Initial status (u_{0i})	31.91	107	295.84	.00
Linear slope (u_{1i})	4.48	107	150.90	.00
Quadratic slope (u_{2i})	.25	107	190.00	.00
Level 1 error (e_{ri})	23.32			

In the between-individual stage, researchers can test whether significant differences exist in growth parameters between distinctive groups of students (e. g., students with LD and those achieving typically). Before relations between growth parameters and correlates (e. g., group membership) are investigated, it is important to examine whether growth parameters are estimated reliably (Willet, 1989a, 1989b).

The reliability of estimated growth parameters in HLM is the ratio of the true parameter variance to the observed variance that consists of the true and error variances (Bryk & Raudenbush, 1987, 1992). If inter-individual differences in growth parameters are mostly due to measurement error, reliable examination of relations between growth parameters and correlates is almost impossible.

Three factors are known to influence the reliability of growth parameters (Willet, 1989b): (a) number of data points, (b) heterogeneity of true growth parameters of individual students, and (c) measurement error. As the number of data points increases, the reliability of growth parameters increases. In addition, the heterogeneity of individual students in true growth parameters is positively related to the reliability of growth parameters. Finally, a smaller measurement error is associated with higher reliability of growth parameters.

HLM-estimated reliability of the initial status parameter of the CBM math data was .60, suggesting 60% of individual differences in the initial status could be attributed to true variations among individuals rather than measurement or sampling errors. In addition, HLM-estimated reliability of the linear growth parameter was .26, whereas the reliability of the quadratic growth parameter was .41. These results indicate that relations between growth rates and correlates (e. g., group membership of LD and non-LD) could be examined somewhat more reliably in the initial status and quadratic growth rate than in the linear growth rate.

Collecting longitudinal data of students with and without LD allows researchers to test the developmental lag hypothesis used earlier as a theoretical basis for selecting a growth model. If the developmental lag hypothesis were to be true, the amounts of math-skill increase should be larger (i. e., higher linear growth rates and smaller deceleration of growth rates) over time for LD students than those for their normal peers. That is, quadratic growth rates for students with LD should be larger than those for students without LD, because the quadratic term shows the amounts of increase in linear growth rates.

To test the developmental lag hypothesis, we used the linear and quadratic growth terms as dependent variables, and group membership (e.g., LD versus non-LD) as an independent variable in the level-two models of HLM. In addition,

background variables affecting students' math achievements (e.g., grade, gender, free/ reduced lunch status) were controlled by including them as covariates, because they are known to have significant relations to individual differences in mathematics achievements. The level-two models used in the analyses were as follows:

$$\pi_{1i} = \beta_{10} + \beta_{11}(LD)_i + \beta_{12}(Grade)_i + \beta_{13}(Gender)_i + \beta_{14}(Lunch)_i + u_{1i},$$

$$\pi_{2i} = \beta_{20} + \beta_{21}(LD)_i + \beta_{22}(Grade)_i + \beta_{23}(Gender)_i + \beta_{24}(Lunch)_i + u_{2i},$$

where π_{1i} and π_{2i} are the linear and quadratic growth rates for individual i, β_{10} and β_{20} the mean linear and quadratic growth rates for third graders who were female, general education students not receiving a free or reduced cost lunch (i.e., independent variables were dummy coded in the analyses), β_{11} and β_{21} the partial regression coefficients indicating the mean group differences in the linear and quadratic growth rates, respectively, between LD and non-LD students with the effects of the other independent variables controlled for, β_{12} and β_{22} the partial regression coefficients indicating the mean group differences between third and fourth graders, β_{13} and β_{23} the partial regression coefficients indicating the mean group differences between males and females, β_{14} and β_{24} the partial regression coefficients indicating the mean group differences between students who received and did not receive a free or reduced cost lunch, and u_{1i} and u_{2i} the random errors. In addition, a group difference in initial status (π_{0i}) between students with and without LD was examined using the same level-two HLM equation described above with the growth parameter of initial status as a dependent variable.

Table 4 shows that a mean initial status for students without LD was higher than that for students with LD by 1.23 problems answered correctly, when the effects of the other covariates were controlled. The mean difference between the two groups, however, was not statistically significant. Mean differences in linear and quadratic growth rates (i. e., 1.80 for linear growth rates and -.42 for quadratic growth rates) were also observed between LD and non-LD groups; however, again, the differences were not statistically significant, when the effects of the other predictors were controlled.

In summary, the results of HLM analyses show that students with LD looked to have lower initial levels of math basic skills at the beginning of the school year, but that they seemed to show similar growth rates to those of their normal peers. When the effects of gender and free or reduced cost lunch status were controlled, HLM-estimated group growth

Table 4. *Differences in Initial Status, Linear Growth, and Quadratic Growth Parameters between Students with and without Learning Disabilities*

Fixed effect	Coefficient	SE	t ratio	df	p value
For initial status					
Intercept (β_{00})	9.86	1.53	6.45	107	.00
LD (β_{01})	-1.23	2.07	0.59	107	.55
Grade (β_{02})	5.37	1.33	4.05	107	.00
Gender (β_{03})	-.26	1.34	.20	107	.85
Lunch (β_{04})	-.86	.68	1.28	107	.20
For linear slope					
Intercept (β_{10})	3.10	.91	3.42	107	.00
LD (β_{11})	1.80	1.24	1.45	107	.15
Grade (β_{12})	.92	.80	1.15	107	.25
Gender (β_{13})	-1.38	.87	1.72	107	.09
Lunch (β_{14})	-.43	.41	1.04	107	.30
For quadratic slope					
Intercept (β_{20})	-.34	.17	2.01	107	.04
LD (β_{21})	-.42	.23	1.83	107	.07
Grade (β_{22})	.08	.15	.56	107	.58
Gender (β_{23})	.32	.15	2.14	107	.03
Lunch (β_{24})	.10	.08	1.35	107	.18

Random effect	SD	Variance	df	χ^2	p value
Initial status (u_{0i})	5.07	25.72	103	249.43	.00
Linear slope (u_{1i})	1.99	3.98	103	138.44	.01
Quadratic slope (u_{2i})	.48	.23	103	178.67	.00
Level 1 error (e_{ti})	4.83	23.29			

curves for third and fourth graders with and without LD were displayed in Figure 2.

In spite of the observed group differences, the analysis results of statistical significance did not support the developmental lag hypothesis. On the contrary, students with LD showed comparable growth trajectories in math skill development with the trajectories of students without LD. The results may be partly due to the fact that the LD participants in the study had been identified as having learning disabilities mainly because of severe reading problems rather than math problems.

Between-Individual Stage: Instructional Factors Facilitating

Academic Growth

One of the major research interests in growth curve analysis is to identify instructional and ecological variables that increase students' growth trajectories. All individual students do not share the same initial status and growth rates because of individual differences in background variables (Bryk & Raudenbush, 1987). For example, one might hypothesize that students who spend more time on homework would show more rapid growth over time than do those who spend less time. Thus, the main focus of the between-individual stage is explaining why some students grow faster than others with student-, instruction-, or ecology-related
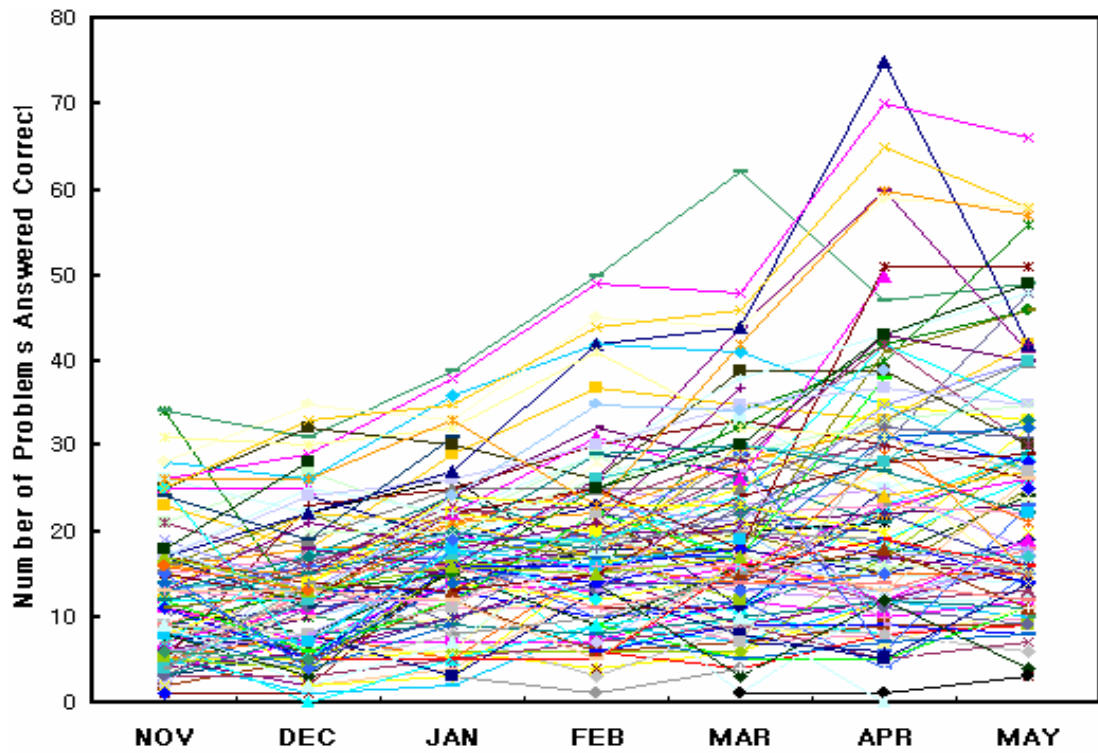
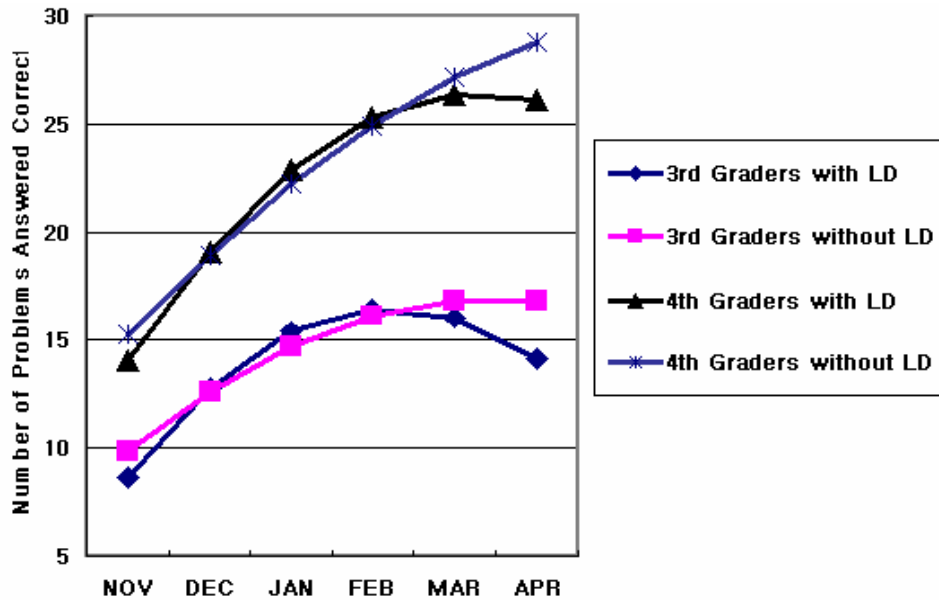*Figure 1.* Individual Students' Growth Curves in CBM Math Data

Figure 2. Estimated Group Growth Curves for Students with and without Learning Disabilities in Grades 3 and 4

Table 5. *Prediction of Inter-individual Differences in Initial Status, and Linear and Quadratic Slopes on Class Participation, Grade, LD Membership, Gender, and Free/ Reduced Lunch Status*

Fixed effect	Coefficient	SE	t ratio	df	p value
For initial status					
Intercept (β_{00})	5.48	2.10	2.61	106	.01
Participation (β_{01})	.17	.06	2.94	106	.00
Grade (β_{02})	5.08	1.29	3.94	106	.00
LD (β_{03})	-1.21	2.01	.61	106	.55
Gender (β_{04})	.21	1.30	.16	106	.87
Lunch (β_{05})	-.55	.66	.82	106	.41
For linear slope					
Intercept (β_{10})	1.05	1.29	.82	106	.41
Participation (β_{11})	.08	.03	2.23	106	.02
Grade (β_{12})	.73	.79	.93	106	.35
LD (β_{13})	1.88	1.22	1.54	106	.12
Gender (β_{14})	-1.16	.80	1.45	106	.15
Lunch (β_{15})	-.32	.41	.79	106	.43
For quadratic slope					
Intercept (β_{20})	-.01	.24	.04	106	.97
Participation (β_{21})	-.01	.01	1.93	106	.05
Grade (β_{22})	.12	.15	.79	106	.43
LD (β_{23})	-.44	.23	1.92	106	.05
Gender (β_{24})	.29	.15	1.90	106	.06
Lunch (β_{25})	.09	.08	1.14	106	.25

Random effect	SD	Variance	df	χ^2	p value
Intercept (u_{0i})	4.78	22.82	102	232.86	.00
Linear slope (u_{1i})	1.90	3.61	102	135.21	.02
Quadratic slope (u_{2i})	.47	.22	102	175.72	.00
Level 1 error (e_{ti})	4.82	23.26	102		

variables as predictors.

For the purposes of demonstration, we examined whether students' active participation in class activities would have a positive relationship with developmental rates of math basic skills. The measure of students' active participation was collected with the Discourse system in the study. Using the system, students typed in their answers and questions during the class activities, and their responses were saved into a file. The number of characters typed in per minute was used for analysis as a measure of students' active participation (see Shin, Deno, Robinson, & Marston, 2000 for more detail).

To investigate the effect of students' active participation on growth rates, we also controlled for the effects of

covariates such as grade, LD membership, gender, and free/ reduced cost lunch status. Level-two models used in these analyses were as follows:

$$\pi_{1i} = \beta_{10} + \beta_{11}(Participation)_i + \beta_{12}(Grade)_i + \beta_{13}(LD)_i + \beta_{14}(Lunch)_i + u_{1i},$$

$$\pi_{2i} = \beta_{20} + \beta_{21}(Participation)_i + \beta_{22}(Grade)_i + \beta_{23}(LD)_i + \beta_{24}(Lunch)_i + u_{2i}$$

where π_{1i} and π_{2i} are the linear and quadratic growth rates for individual i, β_{10} and β_{20} the intercepts of linear and quadratic growth rates, the other β's the partial regression coefficients showing the effects of each independent variable on the linear and quadratic growth rates when the effect of the other predictors are controlled, and u_{1i} and u_{2i} the random errors. In addition, the effects of the independent variables on the initial status were examined with the growth parameter of initial status (π_{0i}) as a dependent variable.

The results of HLM analysis show that higher active participation was significantly related to higher initial levels of math basic skills, when the other factors were controlled (see Table 5). Specifically, students with higher participation rates showed higher achievement by .17 problems than those with lower participation rates. In addition, students who participated more showed higher linear growth rates by .08 problems per month than who did less. Finally, the amounts of decrease of the linear growth rate were larger by .01 problems per month for students with higher participation rates than that for students with lower participation rates.

In summary, the results indicate that students participating more actively in class activities had higher levels of math basic skills and developed their math skills more rapidly than those participating less actively, but that their rates of growth tended to decrease slightly more quickly over time. The results suggest that instructional methods to facilitate students' class participation should be employed more commonly to enhance students' achievement, which is in agreement with the results of previous research on student participation.

Research on Student Growth Using CBM and HLM

As shown with CBM math data, growth patterns, growth rates, and relationship between growth parameters and correlates can be investigated more reliably and accurately by using multiple data points of student performance. In combination with HLM, CBM creates an alternative to the use of difference scores computed with two data points in assessing students' academic growth and its relations to correlates because it produces repeated measures of student performance in a logistically efficient, technically sound way within relatively short time periods. We believe that the value of using CBM and HLM will increase our understanding of developmental characteristics of students with learning difficulties and of educational practices leading to improved learning outcomes over time.

Studies are necessary to examine developmental characteristics of students with learning difficulties in various subject areas, their typical growth rates, and instructional, ecological variables associated with higher rates of academic development. Regarding developmental characteristics, researchers might be interested whether students with learning difficulties would catch up with their normal peers when intensive, effective instructional supports are provided. According to the developmental lag model, students with learning difficulties would show slow growth at the beginning and then would gradually develop academic skills at faster rates over time. As a result, these students might reach the same performance levels with their average-achieving peers due to the effects of intensive, continuous educational services (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). Comparison of individual- and group-growth curves between students with learning difficulties and average-achieving students will enable researchers to examine developmental characteristics of the two types of students.

The study of typical growth rates for students with learning difficulties at each grade level is necessary to establish more sensitive criteria for monitoring these students' progress and evaluating the effectiveness of instructional programs (Deno, Fuchs, Marston, & Shin, in press). Students with learning difficulties develop academic skills at much slower rates than do their peers (Bast & Reitsma, 1998; Francis, 1992). Therefore, using normative information on growth rates estimated mainly from the academic performance of average-achieving students might be too conservative to accurately evaluate the progress of students with learning difficulties and the effectiveness of programs implemented for these students.

Further research is also needed to identify student- and instruction-related factors that facilitate student growth over time. Student-related factors (e. g., active participation in class activities, motivation to learn, and hours of homework) and instruction-related factors (e. g., instructional-planning time, time allocated to instruction, and formative evaluation of student performance) could be associated with inter-individual differences in growth rates. The multi-wave measures of CBM in combination with HLM allow

researchers to investigate such relations between growth rates and student- and instruction-related variables in a more reliable and accurate way.

Furthermore, research on student growth could be extended to program evaluation of different educational services on the basis of changes in growth rates or growth patterns. That is, researchers could evaluate the effectiveness of a special program for students with learning difficulties by examining whether growth trajectories for individuals and groups have changed positively or negatively after implementing the program. Evaluation of program effectiveness can be done more reliably using the growth model approach with multiple data points rather than using simple mean comparisons of static measures.

Examining growth of students with learning difficulties could have important implications for educational decision-making. First, investigating growth patterns of these students across grade levels could help administrators at the school, district, and state levels to determine when resources are most likely to have the greatest impact. For example, suppose that students with severe learning difficulties do not develop their reading fluency until fourth grade. But after fourth grade, these students' growth rates of reading-fluency development are similar to their average-achieving peers (Shin, 1999). In such a case, intensive reading programs might be implemented for students with learning difficulties in the early grades to provide a boost in their reading development.

A second educational implication is that examining typical growth rates of students could provide teachers with guidelines on how much growth they could expect from their students (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). For instance, suppose that students receiving direct-instruction programs in a resource room setting typically show an increase of 3 words-read-correctly per week on CBM reading measures. Resource-room teachers who adopt similar reading programs could expect their students to show at least a three-word increase per week based on such normative information. If a student did not improve his or her reading skills as quickly as expected, the teacher could consider modifying or changing the student's instructional program to increase the rate of student performance.

A final educational implication is that examining relations between school characteristics and student growth could help parents to choose schools that would be most appropriate for their children. In addition, administrators could make more accurate decisions regarding which schools in the district best serve students based on examination of the relationship between school characteristics and student growth.

CBM as a multi-wave growth monitoring system holds promise for using the emerging statistical tool of HLM for assessing the development of the knowledge and skill levels of students with learning difficulties because of its distinctive characteristics. We believe that measuring growth over time better represents the amount of learning of these students, and that CBM and HLM will provide educational researchers with reliable and sensitive tools for assessing growth and making timely educational decisions on individual students.

References

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

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147-158.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newsbury Park, CA: Sage.

Deno, L. S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Deno, L. S., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16.

Deno, L. S., Fuchs, L. S., Marston, D., & Shin, J. (in press). Using Curriculum-Based Measurement to Establish Growth Standards for Students with Learning Disabilities. *School Psychology Review*.

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

Francis, H. (1992). Patterns of reading development in the first school. *British Journal of Educational Psychology, 62*, 225-232.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27-48.

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61- 88). NY: Guilford.

Hertzog, C., & Rovine, M. (1985). Repeated measures analysis of variance in developmental research: Selected issues. *Child Development, 56*, 787-809.

Labouvie, E. W. (1982). The concept of change and regression toward the mean. *Psychological Bulletin, 92*, 251-257.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

McCall, R. B., & Appelbaum, M. (1973). Bias in the repeated measures analysis of variance: Some alternative approaches. *Child Development, 44*, 333-344.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why it is. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18- 78). NY: Guilford.

Marston, D., Deno, S. L., & Tindal, G. (1983). *A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress* (Research Report No. 126). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52*, 266-276.

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Bryk, A. S. (1989). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education, 15*, 423-475.

Rogosa, D. R., & Willet, J. B. (1983). Demonstrating the reliability of the difference scores in the measurement of change. *Journal of Educational Measurement, 20*, 335-343.

Salvia, J., & Ysseldyke, J. E. (1995). *Assessment* (6th Ed.). Boston, MA: Houghton Mifflin Company.

Shaywitz, B. A., & Shaywitz, S. E. (1994). Measurement and analyzing change. In Lyon, G. R. (Ed.), *Frames of reference for the assessment of learning disabilities* (pp. 59-67). Baltimore, MD: Brookes.

Shin, J. (1999). *Reading-skill development and instructional practices facilitating reading growth for students with and without learning disabilities: A one-year longitudinal study*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Shin, J., Deno, S. L., & Espin, C. A. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education, 34*, 164-172.

Shin, J., Deno, S. L., Robinson, S. L., & Marston, D. (2000). Predicting classroom achievement from active responding on a computer-based groupware system. *Remedial and Special Education, 21,* 53-60.

Stanovich, K. E., Nathan, R. G., & Zolman, J. E. (1988). The developmental lag hypothesis in reading: Longitudinal and matched reading-level comparisons. *Child Development, 59*, 71-86.

Willet, J. B. (1989a). Questions and answers in the measurement of change. *Review of Research in Education, 15*, 345-421.

Willet, J. B. (1989b). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49*, 587-603.

Willet, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381.

Zimmerman, D. W., & Williams, R. H. (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement, 19*, 149-154.