Exploring the Relationship Between Static and Dynamic Vertical Scaling from Cross-Sectional and Longitudinal Design Perspectives

Shudong Wang
Educational Testing Service

Hong Jiao
University of Maryland

Yanming Jiang
Educational Testing Service

Paper presented at the annual meeting of the National Council on Measurement in Education (NCME)

April 13-17, 2009, San Diego, CA.

Abstract

The concept of dynamic vertical scaling (DVS) from longitudinal point of view has been proposed as comparing to traditional vertical scaling or static vertical scaling (SVS) from cross sectional perspective. The effects of differences between DVS and SVS on large-scale student achievements have been investigated. The potential application of DVS on creating or explaining students growths was explored. In general, DVS can account for many factors that could affect students growth where SVS has limitations on its cross-sectional design for applications unless strong assumptions has to be made.

Introduction

The purpose of vertical scaling is to measure student growth in achievement by linking tests that (a) measure a similar construct across grades and (b) differ in difficulty across grades. Though vertical scales of large-scale commercial achievement tests and state assessments are created with different scaling designs, they all employ cross-sectional data collection designs. Later when these vertical scales are used in practice, they are interpreted from a longitudinal perspective. Due to a variety of reasons, the growth of students could change from year to year. For example, establishing a vertical scale in 2005 could be different from establishing a vertical scale in 2006 due to policy, curriculum, schooling, and social economics factors changes. While conducting vertical scaling every year may closely match student real growth, practical limitations, such as costs and collecting the scaling data, can make it impossible.

Traditional vertical scaling is called static vertical scaling (SVS) because once the vertical scale is created; it is assumed that there is no change in student growth across many years from grade to grade. Dynamic vertical scaling (DVS), on the other hand, treats vertical scaling constants as a function of time. The major application steps of SVS (see Figure 1) in the real world are:

1. Create the vertical scale for Form A at year one based on samples of students across different grades using one of the vertical scaling designs.
2. Apply vertical scaling constants obtained with Form A across grades to measure students growth for many years down the road.
3. When there is a need to create a new Form B, equate Form A and Form B horizontally at each grade. Apply the same vertical scaling constants from Form A across grades plus equating constant between Forms A and B to Form B to measure students growth using Form B.

By applying the vertical scaling from one year to others, a strong assumption has been made: student achievement growth, i.e., the mean differences between grades, are constant across cohort groups at different time points. In reality, however, the growth may vary from cohort to cohort and from year to year (see Figure 2).

The No Child Left Behind Act of 2001 (NCLB) requires the reporting of adequate yearly progress (AYP) for student performance within selected subject areas. Although, a vertical scale is not necessary for AYP, in a few states the properties of a vertical scale are used as part of growth models in AYP. Many standardized large-scale state criterion-referenced tests and commercial norm-referenced achievement tests such as *The Iowa Tests of Basic Skills* (ITBS, Hoover, Dunbar, & Frisbie, 2003), *TerraNova* (CTB/McGraw-Hill, 1999), and *Stanford Achievement Test* (Harcourt Assessment, 2004) all report scores on a vertical scale that allows assessment of student group growth trends and individual growth in achievement. Typically, for commercial achievement tests, multiple subject tests are constructed by using either classical test theory or item response theory (IRT) to assess student performance at each grade, and test scores are linked to a vertical scale because a well-constructed vertical scale connects forms across different grades and can be used to illustrate patterns of growth over time and model changes in student achievement (Kolen & Brennan, 2004).

In spite of the many advantages using a vertical scale in large-scale assessment, vertical scaling has been considered as a very challenging psychometric procedure for many years (Feuer, Holland, Green, Bertenthal & Hemphill, 1999). Some major challenges include that different types of errors can be introduced into a vertical scale because of limitation in method and design used to create the scale, therefore, the scale may be flawed (Camilli, 1999; Doran & Cohen, 2005; Doran & Jiang, 2006; Lissitz & Huynh, 2003& 2005; Schafer, 2006). Some possible sources of error or major criticism of vertical scale include (1) stages of development represent different latent traits, not levels of attainment of a single trait; (2) constructs exhibit differences in dimensionality at different time points; and (3) violations of unidimensionality can prevent accurate measurement of ability. (Haertel, 1991; Martineau, 2006; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). However, one more source of error which is often ignored about vertical scaling is the difference between static (cross-sectional) and dynamic (longitudinal) vertical scaling designs.

As described earlier, for static vertical scaling, the fixed vertical scaling constants created at one time point are applied to different cohort groups across time, i.e., making the assumption that the vertical scaling is time invariant. On the other hand, with dynamic scaling the vertical scaling constants vary with time. One of the strengths of DVS is that it can reveal the impact of the particular circumstances in which students grow and reflect student real longitudinal development. For example, Table 1 shows an example of a vertical scale across grades for different cohorts (C_1 to C_{10}). If a vertical scale is created at year Y_1 (base year) based on cross-sectional sample data of C_4 to C_{10}, then for next 4 years, the same growth pattern between grades 4 to 8 will apply for years Y_2 to Y_5, despite of the fact that the nature of cross-sectional sample data are very different (Y_2 is C_3 to C_9, Y_3 is C_2 to C_8, …etc.). The differences between SVS and DVS are shown in Figures 1 and 2 based on fabricated examples; in these examples, the vertical constant is assumed to be in logit scale based on the IRT Rasch model.

Conventional SVS methods either avoid the issue of the vertical scale that is created based on different cross-section data or ignore the fact that vertical constants may be a function of time. Ignoring the difference among cross-sectional data across years may result in errors in growth estimation which may lead to misleading conclusions of student growth. DVS based on a longitudinal growth model that reflects longitudinal change in populations across time combined with SVS method allows the potential capability to address the issue directly by modifying base year scaling constants using a longitudinal model based on accumulated longitudinal data over a period of time. Because learning by definition is a change phenomenon and DVS focus on this phenomenon, a longitudinal design represents it well.

The purpose of this study is to demonstrate differences between static and dynamic vertical scales inherent in data collection designs and applications. The current study explores the relationship between SVS and DVS using simulation approaches and provides an understanding on how differently they reflect student true achievement growth.

Methods

In educational assessment, estimation errors mainly come from estimation of person effects, item effects, or a combination of the two. In vertical scaling, another source of error may come from changes in the population across years. In the SVS approach, population variation is not taken into consideration. Thus, this study focuses on sample variations over time in addition to the nested structure in the sampling process. Because true vertical scaling constants are unknown for both DVS and SVS, only estimates of them may be obtained. The best way to evaluate estimation error is to use simulation methods in which true parameters are known.

In practice, only a single observed linking constant, $\hat{L}$, an estimate of the true unknown linking parameter *L,* is obtained and it can be expressed as a liner combination of *L* plus an error term. In theory, if it is possible to conduct vertical scaling repeatedly either at the same time (using different random samples) or at different time points (years), there is variance of L or $var(L)$. The deviation of $\hat{L}$ from L could come from both random and system errors. For SVS using cross-sectional data at the same time point, the source of error comes from inter-individual difference, while for DVS using longitudinal data at different time points, the source of error comes from both inter- and intra-individual differences. The assumption of the traditional SVS application has a shortcoming: by ignoring the difference between cohorts as a function of time implies, SVS treats intra-individual (change within individual) differences as constants across time (here time is year, see Table 1) and this deficiency is inherant in all cross-sectional designs. With longitudinal designs, on the other hand, DVS tries to model intra-individual differences differently across time. Besides variation in how individual differences are treated, there are also additional three sources of variation associated with $\hat{L}$ for within individual differences that DVS could model: (1) measurement error; (2) unobserved covariates; (3) serial correlation (Diggle, Liang, & Zeger, 2002). In practice, because vertical scaling is only conducted once, the variance of L is zero. By using simulation methods (Monte Carlo), the linking variance on vertical scaling constants can be quantified. The departures of $\hat{L}$ from L can be modeled by replicating various simulation conditions.

A. Dynamic Vertical Scaling

Longitudinal ability θ in DVS are created by using both linear and nonlinear mixed models (LMM, Goldstein, 1995; Molenburg & Verbeke, 2005; Raudenbush & Bryk, 2002) to generate correlated multivariate longitudinal "true θ" over time for a student in a particular cohort across grades. Student ability θ across grades can be expressed (in the logit scale) in the Rasch model as an input for the longitudinal model. Before further discussion, two terms should be clearly defined: 1) between-cohort differences refer to the differences between grades evaluated at different times; 2) inter-individual differences refer to the difference between an individual within a particular cohort (grade) and does not mean the difference across cohorts. For this study, individual student growth is measured by both linear and non-linear (quadratic form) LLM models, see Appendix A. LMM has many advantages for longitudinal data that are related to this study: LMM can model individual change over time, model between-and within-individual variations, model the covariance structure of the repeated measures, model non-linear patterns (polynomial or spline), and model higher level data structures (such as student nested within school). The general equation for LMM (Little, Milliken, Stroup, & Wolfinger, 1996) is

$$\theta_i = \mathbf{X}_i\mathbf{\Gamma} + \mathbf{Z}_i\mathbf{U}_i + \mathbf{R}_i \tag{1}$$

For example, nonlinear growth for a cohort can be modeled as:

Level-1 (within Individual)

Quadratic growth $\theta_{it} = \pi_{0i} + \pi_{1i}\cdot\text{Year}_{it} + \pi_{2i}\cdot\text{Year}_{it}^2 + e_{it}$. (2)

Level-2 (Between Individual)

Intercept $\pi_{0i} = \beta_{00} + \text{Covariate}_0 + u_{0i}$, (2a)

Linear slope $\pi_{1i} = \beta_{10} + \text{Covariate}_1 + u_{1i}$, (2b)

Quadratic slope $\pi_{2i} = \beta_{20} + u_{2i}$, (2c)

where in the level 1 model, θ_{it} is the ability of student i on the logit scale at time point t (=year-1), t= 0, 1, 2, … T. Equations (2) can be expressed in matrix format as (without concerning covariate terms):

$$\begin{pmatrix}\theta_{i0}\\ \theta_{i1}\\ \vdots\\ \vdots\\ \theta_{iT}\end{pmatrix} = \begin{pmatrix}1 & \text{Year}_{i0} & \text{Year}_{i0}^2\\ 1 & \text{Year}_{i1} & \text{Year}_{i1}^2\\ \vdots & \vdots & \vdots\\ \vdots & \vdots & \vdots\\ 1 & \text{Year}_{iT} & \text{Year}_{iT}^2\end{pmatrix}\begin{pmatrix}\beta_{00}\\ \beta_{10}\\ \beta_{20}\end{pmatrix} + \begin{pmatrix}1 & \text{Year}_{i0} & \text{Year}_{i0}^2\\ 1 & \text{Year}_{i1} & \text{Year}_{i1}^2\\ \vdots & \vdots & \vdots\\ \vdots & \vdots & \vdots\\ 1 & \text{Year}_{iT} & \text{Year}_{iT}^2\end{pmatrix}\begin{pmatrix}u_{0i}\\ u_{1i}\\ u_{2i}\end{pmatrix} + \begin{pmatrix}e_{i0}\\ e_{i1}\\ \vdots\\ \vdots\\ e_{iT}\end{pmatrix} \tag{3}$$

In this study, only the random intercept and random slope models are examined, so the quadratic slope was fixed and u_{2i}= 0. The distributions of the random effect of the nonlinear model used in this study (see equation A2h in Appendix A) are

$$\begin{bmatrix}u_{0i}\\ u_{1i}\\ u_{2i}\\ e_{it}\end{bmatrix} \sim N\left[\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}\tau_{00}^2 & \tau_{10}^2 & \tau_{20}^2 & 0\\ \tau_{10}^2 & \tau_{11}^2 & \tau_{12}^2 & 0\\ \tau_{20}^2 & \tau_{12}^2 & \tau_{22}^2 & 0\\ 0 & 0 & 0 & \sigma^2\end{pmatrix}\right] = N\left[\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.4 & 0.05 & 0\\ 0.4 & 1 & 0.05 & 0\\ 0.05 & 0.05 & 0.03 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\right] \tag{4}$$

In order to match the SVS simulation design, a common-person design is used to generate longitudinal data, i.e., at each year, students took on-grade and below-grade tests with 40 items in each. Simulated samples are drawn from a multivariate normal distribution with specified mean vector and variance-covariance matrix for each cohort. In general, the correlation of θ between adjacent years is around 0.8 (Lockwood & McCaffrey, 2007) for achievement tests, and as the time lag increases, the correlation decreases. Because multiple cohorts' data are needed, multiple samples are generated using different sets of parameter values.

For all cohorts, the unconditional means (β_{00}) matches the means of SVS for both on-grade and below-grade results. A sample size of 2000 per cohort is used.

B. Static Vertical Scaling

Linear growth can be modeled as in equation (A1) in Appendix A. The attempt here is to incorporate SVS into a DVS frame work and, as a matter of fact, the SVS is a special case of the longitudinal model. At the base year (year = 1), student achievement can be expressed as:

$$\theta_{it} = (\beta_{00} + u_{0i}) + (\beta_{10} + u_{1i}) \cdot \mathrm{Year}_{it} + e_{it} = \beta_{00} + u_{0i} + e_{it} \quad (5)$$

where year= $t + 1$, (t=0, 1, 2, …T) presents time and i (i=1,2,3,…N_p) presents person. For a particular cohort at a particular time point, the SVS model can be expressed as an unconditional mean model or the null model (in both the linear and nonlinear cases), as in equation (5) except there is no modeling of intra-individual variation. Because of this, the level-1 variance is zero and the student score for grades x and y can be expressed as:

$$\theta_{itx} = \beta_{00x} + u_{0ix} \quad (6)$$
$$\theta_{ity} = \beta_{00y} + u_{0iy} \quad (7)$$

The average difference between grade x and y is the vertical scaling constant, i.e.:

$$\text{Scaling-Constant} = \beta_{00x} - \beta_{00y} \quad (8)$$

The conventional SVS method is used to create reference vertical scale (base year scale) in this study. A common-person scaling design is used in which both below-grade and on-grade items are used to link adjacent grades as shown in Table 2a. As in Table 1, to simulate multiple "true score" growth patterns (C_1 - C_{10} at Y_1 with a cross-sectional design, which will apply to the rest of the years), student abilities θ are generated from a multivariate normal distribution (N(M,SD)). 2000 students for each grade and 40 items for each grade are also simulated, and the mean Rasch difficulty of items are listed in Table 3. For other years (not base years), a first order autoregressive process (AR1) is used to model serial correlation (part C in Appendix A). For example, parameters for linear growth of both intercept (Part A in Appendix A) and intercept plus slope models for on-grade can be generated as:

$$\mathbf{Y} = \begin{vmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \\ \theta_9 \\ \theta_{10} \end{vmatrix} \sim MVN \left| \begin{pmatrix} -2 \\ -1.6 \\ -1.2 \\ -0.8 \\ -0.4 \\ 0 \\ 0.4 \\ 0.8 \\ 1.2 \\ 1.6 \end{pmatrix}, \begin{pmatrix} 1 & & & & & & & & & \\ 0.80 & 1 & & & & & & & & \\ 0.64 & 0.80 & 1 & & & & & & & \\ 0.51 & 0.64 & 0.80 & 1 & & & & & & \\ 0.41 & 0.51 & 0.64 & 0.80 & 1 & & & & & \\ 0.33 & 0.41 & 0.51 & 0.60 & 0.80 & 1 & & & & \\ 0.26 & 0.33 & 0.41 & 0.51 & 0.64 & 0.80 & 1 & & & \\ 0.21 & 0.26 & 0.33 & 0.41 & 0.51 & 0.64 & 0.80 & 1 & & \\ 0.17 & 0.21 & 0.26 & 0.33 & 0.41 & 0.51 & 0.64 & 0.80 & 1 & \\ 0.13 & 0.17 & 0.21 & 0.26 & 0.33 & 0.41 & 0.51 & 0.64 & 0.80 & 1 \end{pmatrix} \right| \quad (9)$$

The difference between cross-sectional and longitudinal data resembles the difference between univariate and multivariate data. Linear and non-linear growth parameters are shown in Figures B1 and B2 in Appendix B. After generating the longitudinal data, only the base year was used for the SVS, because that model assumes the cross-sectional data apply to multiple years. The mean and variance parameters of the linear intercept model used in this study (see equation A1 in Appendix A) are

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ e_{it} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{10}^2 & 0 \\ \tau_{10}^2 & \tau_1^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}\right] = N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.4 & 0 \\ 0.4 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right] \quad (10)$$

Because the purpose of this study is to explore the accuracy of vertical scaling across years using DVS versus time-invariant SVS, in order to see the effect of different types of growth, the artificially induced constant “noise” in growth that show non-constant differences of growth among cohorts are added to two cohorts (C1 and C6, grade 1 and grade 6 at the base year) throughout the years (year 1-5). For both linear and nonlinear model, the covariates are added into level-2 equations that present inter-individual differences (see equations A1a-A1b and A2a-A2c in Appendix A). Two simulation approaches are used to explore the effect of DVS.

Approach I

First approach is to apply common practices method on different types of growth. In this approach, the recovered growth patterns are compared with the simulated true patterns in the following section. The major steps of the Monte Carlo simulation of first approach are:

1. Generate both linear and nonlinear true growth (See Figures B1 and B2 in Appendix B) using the unconditional longitudinal model and multivariate distributions with known true parameters for base year. The unconditional model is also called the null model, which mirrors cross-sectional data. The scaling design is a common-person design.
2. Create SVS by using separate calibrations applying the Rasch model by grade and connecting the vertical scale using 40 common items with the Stocking-Lord (Stocking & Lord, 1983) linking procedure, in which both on- and below-grade tests taken by the same students are linked (see Table 2a).
3. Using the scale obtained from SVS to generate linear and nonlinear true growth with or without "noise" using both random intercept and random slope longitudinal models for both on- and below grades. Both random intercept and random slope longitudinal models introduce time variant errors where SVS does not have these.
4. Conduct separate calibration for each grade.
5. Create DVS by linking each year scale by grade to the base year vertical scale using 10 common-items equating design with the Stocking-Lord linking procedure (see Table 2b).
6. Winsteps (Linacre, 2008) software was used to conduct item calibration by grade and IceDog software (Robin, Holland, & Hemat, 2006) was used to estimate the Stocking-Lord transformation function constants, intercept B and slope A.

There are a total 2 growth patterns (Linear model & nonlinear model) x 3 models (null model, random intercept model, and random intercept and slope model) x 2 cohort changes (non-noise and noise) – 2 (null model has no noise) = 10 simulation conditions. The bias, SE, and RMSE are used to evaluate how well true parameters are recovered for each of 6 simulation conditions. These formulas are,

$$Bias(\hat{\theta}_g) = \frac{1}{N_R}\sum_{r=1}^{N_R}\left(\frac{\sum_{i=1}^{N_p}\hat{\theta}_{gri}}{N_p} - \overline{\theta}_g\right), \tag{11}$$

$$SE(\hat{\theta}_g) = \sqrt{\frac{1}{N_R}\sum_{r=1}^{N_R}\left(\frac{\sum_{i=1}^{N_p}\hat{\theta}_{gri}}{N_p} - \frac{1}{N_R}\frac{1}{N_p}\sum_{r=1}^{N_R}\sum_{i=1}^{N_p}\hat{\theta}_{gri}\right)^2}, \tag{12}$$

$$RMSE(\hat{\theta}_g) = \sqrt{\frac{1}{N_R}\sum_{r=1}^{N_R}\left(\frac{\sum_{i=1}^{N_p}\hat{\theta}_{gri}}{N_p} - \overline{\theta}_g\right)^2}. \tag{13}$$

where *i, g,* and *r* represent individual, grade, and replication, respectively. $\hat{\theta}_{gri}$ is the estimated person parameter for grade *g*, replication *r*, and person *i*. $\overline{\theta}_g$ is the mean of the generated true students' abilities in grade *g* used in the null models. N_p is the number of simulated examinees and N_R is the number of replications of the simulation. There are a total of 50 replications used in this study. Different random numbers are used as seeds for each of the 50 replications. Based on a past research suggestion (Harwell, Stone, Hsu, & Kirisci, 1996), descriptive procedures are used to summarize the simulation results. All estimates for the 10 different simulation conditions are compared to SVS.

Approach II

Second approach is to conduct vertical scaling at each year by using the same vertical scaling design as SVS used for the base year and the vertical scaling constants at each year are then compared for different growth patterns. If there are different vertical growths in each cross-sectional data for certain year, the difference of cross-sectional growth should be reflected by vertical scaling constant by that year. After all, DVS is multiple SVS at different time point. The major steps of the Monte Carlo simulation of second approach are:

1. Generate both linear and nonlinear true growth (See Figures B1 and B2 in Appendix B) using the unconditional longitudinal model and multivariate distributions with known true parameters for base year. The unconditional model is also called the null model, which mirrors cross-sectional data. The scaling design is a common-person design.
2. Create SVS by using separate calibrations applying the Rasch model by grade and connecting the vertical scale using 40 common items with the Stocking-Lord (Stocking & Lord, 1983) linking procedure, in which both on- and below-grade tests taken by the same students are linked (see Table 2a).
3. Using the scale obtained from SVS to generate linear and nonlinear true growth with or without "noise" using both random intercept and random slope longitudinal models for both on- and below grades for different cohorts from year 2 to year 5. Both random intercept and random slope longitudinal models introduce time variant errors that SVS does not have.
4. Conduct separate SVS at each year (year 2 to year 5) across grades based on step 2.
5. Winsteps (Linacre, 2008) software was used to conduct item calibration by grade and IceDog software (Robin, Holland, & Hemat, 2006) was used to estimate the Stocking-Lord transformation function constants, intercept B and slope A.

Results

A. Results of Approach I

The major goal of this study is to examine the relationship between SVS and DVS as true growth changes with time. If vertical scaling treats the dynamic growth as static growth, then how much distortion will SVS produce? The effect of the distorted SVS is examined in terms of bias, SE, and RMSE.

A1. Linear growth

Tables 4 and 5 lists the bias, SE, and RMSE of 6 different linear model conditions without and with noise. Figures 3 and 4 also show the bias, SE, and RMSE of linear models without and with noise across years and grades. It is clear that on average, bias, RMSE, and SE increase as the model becomes more complex. In general, the null model has less bias, SE, and RMSE than the random intercept model, and the intercept model has less bias, SE, and RMSE than the random intercept + slope model. Models with noise have larger bias, SE, and RMSE than these models without noise. This is no surprise because the SVS method only accounts for cross-sectional data as a one-time estimate of true growth. When true growth includes both inter- and intra-individual variation, SVS cannot capture multiple sources of variation. The presented tables and figures also show that the more distant from the base year, the larger the bias, SE, and RMSE become. In particular, for all three models the deviations of SVS from the true model increase with time (year) (Figures 3 and 4). This result implies that once the vertical scale was created, the longer time it is used, the larger chance it misrepresents the true growth.

A further observation is that, in general, as grade increases, the bias, SE, and RMSE increase. This result is an artifact of the vertical scaling design, which used a chain linking design in which grade 1 was chosen as the base and each grade was linked to the next lower grade; given this linking design, the errors are accumulated going up the grades. This trend is especially evident for the SE when years are close to the base year. If the linking design had chosen grade 10 as the base year, then the opposite effect would have been observed.

A2. Nonlinear growth

Tables 6 and 7 presents the bias, SE, and RMSE of 6 different nonlinear model without and with noise. Figures 5 and 6 demonstrate the bias, SE, and RMSE of nonlinear models across years and grades. Similar to the linear model results, on average, bias, RMSE and SE increase as models change from null to intercept, and from intercept to intercept + slope. The models without noise have less bias, SE, and RMSE. The reason for this trend is the same as the linear cases, in that the SVS method created for cross-sectional data design cannot account for the true growth that includes across-year change. When true growth includes both within- and between-year variations, SVS cannot accurately reflect this type of growth. Tables 4 and 5 and Figures 2 and 4 also show that for years further apart from the base year, the larger the bias, SE, and RMSE. In particular, the variations for all three models show the increase in accuracy when the year increases. This result indicates that the older the vertical scale becomes, the larger the chance it misrepresents true growth.

Other evidence shows, in general, as grade increases, the bias, SE, and RMSE increase. As described earlier, this is an artifact of the chain vertical scaling design used.

In general, for both linear and nonlinear models, the chance of errors increases when: 1) the model becomes more complex or realistic and 2) year increases or is farther apart from the base year when vertical scale was created.

B. Results of Approach II

Table 8 list simulated vertical scaling constants (VSC) under different models with and without noise and Figures 7 and 8 (Null models figures are based on assumption: VSC is constant across years) depict these VSCs for different models without and with noise for the second approach. It is clearly that the VSC is function of time (year) and assumption of constant vertical scaling doesn’t hold for any of the given simulated conditions. In general, as growth model become more complex from null model to intercept plus slope model without and with noise, the deviations of VSC from SVC increases, in particular, the noise has an important impact on the VSC, which implies that ignoring the difference between cohorts as SVS does distort the VSC because SVS doesn’t account for student’s longitudinal growth.

Discussion and Summary

Many state and standardized large-scale achievement tests report scores on a vertical scale that allows assessment of student group trends and individual growth in achievement. However, developing a valid vertical scale is a time consuming, expensive and complicated process so that test developers cannot update the vertical scale frequently. The vertical scaling design, growth pattern, and dimensionality of the data are some of the most important considerations in creating the scale. Because of major limitation on time and cost to create vertical scales, common practice is to make the strong assumption that once a vertical scale is created at a particular time point by using a particular sample, applying a particular design and model, then one can treat the estimates of grade-to-grade growth constants as invariant to reflect student growth for all assessments tied to that vertical scale. However, the nature of student growths is a very complex social and educational phenomenon. Ignoring the complexity of growth phenomena can distort student achievement results. This paper focuses on how much difference SVS and DVS can make by simulating different growth patterns and using different models that try to mirror as closely as possible actual student growth (such as variation within student scores). Given the reality of creating a vertical scale, this paper attempted to demonstrate the distorted dynamic growth pattern that can occur when using the static modeling approach and explore the theoretical possibility of using longitudinal modeling to improve the quality of the vertical scale. For example, each year, some states and testing vendors anticipate the so called "unexpected changes" of student's achievements when comparing to previous year (see Figure 9) results, i.e., a large decrease or increase in the proportion of student passing certain performance standard. One of the most difficulty tasks for state and testing vendor in this situation is to explain the reason(s) for such phenomenon. This study proposes a concept of DVS and underscores the importance of longitudinal points of view of student development. It provides an alternative way to develop a valid vertical scale, to explain the "unexpected changes", and possibly to strengthen the defensibility of using vertical scales.

The limitations of the proposed DVS method will be the increase of the complexity in the construction process and the interpretations of the vertical scale. DVS would complicate the test design by including the use and analysis of cross-grade linking items every year. In addition, the DVS would complicate the interpretation of scale scores, because the scale score system would change every year. What the DVS does contribute is the possibility of more accurate comparison of amounts growth at different parts of a scale at one point of time (e.g., longitudinal growth from grade 3 to grade 4 compared to longitudinal growth from grade 4 to grade 5 for one particular pair of years).

This paper only puts forward a hypothesis. For example, no missing cases across time are assumed in the longitudinal model. Whether the real world changes in a manner like that simulated in this study also awaits further investigation. The next step of this line of research is to find longitudinal data sets which may fall close to the design simulated in this study and use the proposed approach to examine the utility of DVS in analyzing real data.

Appendix A

Longitudinal Linear and Nonlinear Models Used to Generate Data

A. Linear Model

Level-1 (Within Individual)

Linear growth $\theta_{it} = \pi_{0i} + \pi_{1i}\text{Year}_{it} + e_{it}$. (A1)

Level-2 (Between Individual)

Intercept $\pi_{0i} = \beta_{00} + \text{Covariate}_0 + u_{0i}$, (A1a)

Linear slope $\pi_{1i} = \beta_{10} + \text{Covariate}_1 + u_{1i}$, (A1b)

In the level 1 model, θ_{it} is the score in logits at each time point, i indicates the student and t is "time" (Year = t + 1), often coded as t = 0, 1, 2, … T. There is no second level covariate for most of grade except for grade 1 and 6 in this study.

θ_{it} = individual score in logits
π_{0i} = individual true score at time 0
π_{1i} = individual true linear slope (linear change of rate)
β_{00} = the average overall true initial score at time 0
β_{10} = the average overall true linear change between individual
Covariate_0 (noise)
= 0 for growth patterns that are the same as SVS
= 0.1 for growth patterns that are different as SVS for cohort at grade 1 (Table 1)
= -0.1 for growth patterns that are different as SVS for cohort at grade 6 (Table 1)
Covariate_1 (noise)
= 0 for growth patterns that are the same as SVS
= 0.05 for growth patterns that are different as SVS for cohort at grade 1 (Table 1)
= -0.05 for growth patterns that are different as SVS for cohort at grade 6 (Table 1)

e_{it} = The random error with the t^{th} time point in the i^{th} individual.
u_{0i}= The random effect for intercepts
u_{1i} = The random effect for linear slopes

The level-1 error term follows a normal distribution, $e_{ti} \sim N(0, \sigma^2)$, where, the common variance σ^2 is determined by the reliability average across the level-1 coefficients (cf. Raudenbush & Bryk, 2002). $e_{it} \sim N(0, \sigma^2)$ it is assumed to be independent of level-2 random effects (u_{0i}, and u_{1i},), that is $\text{cov}(e_{it}, u_i) = 0$.

The level-2 error terms also follows normal distributions:
$u_{0i} \sim N(0, \tau_0^2)$, τ_0 is the variance of level 2 residuals u_{0i} from predicting the level 1 intercept (π_{0i})
$u_{1i} \sim N(0, \tau_1^2)$, τ_1 is the variance of level 2 residuals u_{1i} from predicting the level 1 slope (π_{1i})

$\mathrm{cov}(u_{0i}, u_{1i}) = \mathrm{cov}(\pi_{0i}, \pi_{1i}) = \tau_{10}$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ e_{it} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{10}^2 & 0 \\ \tau_{10}^2 & \tau_1^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}\right] \quad \text{(A1c)}$$

Together for linear growth (without Covariates):

$$\theta_{it} = (\beta_{00} + u_{0i}) + (\beta_{10} + u_{1i}) \cdot \mathrm{Year}_{it} + e_{it} = (\beta_{00} + \beta_{10} * \mathrm{Year}_{it}) + (u_{0i} + u_{1i} * \mathrm{Year}_{it} + e_{it})$$
$$= (\text{fixed}) + (\text{random}) \quad \text{(A1d)}$$

$$\begin{pmatrix} \theta_{i0} \\ \theta_{i1} \\ : \\ : \\ \theta_{iT} \end{pmatrix} = \begin{pmatrix} 1 & \mathrm{Year}_{i0} \\ 1 & \mathrm{Year}_{i1} \\ : & : \\ : & : \\ 1 & \mathrm{Year}_{iT} \end{pmatrix} \begin{pmatrix} \beta_{00} \\ \beta_{10} \end{pmatrix} + \begin{pmatrix} 1 & \mathrm{Year}_{i0} \\ 1 & \mathrm{Year}_{i1} \\ : & : \\ : & : \\ 1 & \mathrm{Year}_{iT} \end{pmatrix} \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ : \\ : \\ e_{in} \end{pmatrix} \quad \text{(A1e)}$$

$$\boldsymbol{\theta}_i = \mathbf{X}_i\boldsymbol{\Gamma} + \mathbf{Z}_i\mathbf{U}_i + \mathbf{E}_i \quad \text{(A1f)}$$

where i =1,….N (number of person) and n = number of time points for individual i.
The marginal model (without Covariates) can be expressed:

$$\mathrm{Var}(\theta_{it}) = E[(\theta_{it} - E(\theta_{it}))^2] = \tau_0^2 + 2\,\tau_{10}\,\mathrm{Year}_{it} + \tau_1^2\,\mathrm{Year}_{it}^2 + \sigma^2 \quad \text{(A1g)}$$
$$\theta_{it} \sim N((\beta_{00} + \beta_{10} * \mathrm{Year}_{it}),\ (\tau_0^2 + 2\,\tau_{10}\,\mathrm{Year}_{it} + \tau_1^2\,\mathrm{Year}_{it}^2 + \sigma^2)) \quad \text{(A1h)}$$

B. Nonlinear Model

Level-1 (Within Individual)

Quadrature growth $\theta_{it} = \pi_{0i} + \pi_{1i} \cdot \mathrm{Year}_{it} + \pi_{2i} \cdot \mathrm{Year}_{it}^2 + e_{it}$. (A2)

Level-2 (Between Individual)

Intercept $\pi_{0i} = \beta_{00} + \mathrm{Covariate}_0 + u_{0i}$, (A2a)

Linear slope $\pi_{1i} = \beta_{10} + \mathrm{Covariate}_1 + u_{1i}$, (A2b)

Quadrature slope $\pi_{2i} = \beta_{20} + u_{2i}$, (A2c)

Together for quadrature growth (without covariates):

$$\theta_{it} = (\beta_{00} + u_{0i}) + (\beta_{10} + u_{1i}) \cdot \text{Year}_{it} + (\beta_{20} + u_{2i}) * \text{Year}_{it}^2 + e_{it}.$$
$$= \beta_{00} + \beta_{10} * \text{Year}_{it} + \beta_{20} * \text{Year}_{it}^2 + u_{0i} + u_{1i} * \text{Year}_{it} + u_{1i} * \text{Year}_{it}^2 + e_{it}. \quad \text{(A2d)}$$

In the level 1 model, θ_{it} is achievement score in logits at each time point, i indicates the student and t is "time" (Year = time + 1), often coded as t= 0, 1, 2, … T. There is no second level covariate for most of grade except for grade 1 and 6 in this study.

θ_{it} = individual score in logit
π_{0i} = individual true score at time 0
π_{1i} = individual true linear slope (linear change of rate)
π_{2i} = individual true quadrature slope (quadrature change of rate)
β_{00} = the average overall true initial score at time 0
β_{10} = the average overall true linear change between individual
β_{20} = the average overall true quadrature change between individual
Covariate_0 (noise)
= 0 for growth patterns that are the same as SVS
= 0.1 for growth patterns that are different as SVS for cohort at grade 1 (Table 1)
= -0.1 for growth patterns that are different as SVS for cohort at grade 6 (Table 1)
Covariate_1 (noise)
= 0 for growth patterns that are the same as SVS
= 0.05 for growth patterns that are different as SVS for cohort at grade 1 (Table 1)
= -0.05 for growth patterns that are different as SVS for cohort at grade 6 (Table 1)
e_{it} : The random error with the t^{th} time point in the i^{th} individual.
u_{0i}: The random effect for intercepts
u_{1i}: The random effect for linear slopes
u_{2i}: The random effect for quadrature slopes

$e_{it} \sim N(0, \sigma^2)$ it is assumed to be independent of level-2 random effects (u_{0i}, u_{1i}, and u_{2i}), that is $\text{cov}(e_{ti}, u_i) = 0$.
$u_{0i} \sim N(0, \tau_{00}^2)$, τ_{00} is the var of level 2 residuals u_{0i} from predicting the level 1 intercept (π_{0i})
$u_{1i} \sim N(0, \tau_{11}^2)$, τ_{11} is the var of level 2 residuals u_{1i} from predicting the level 1 linear slope (π_{1i})
$u_{2i} \sim N(0, \tau_{22}^2)$, τ_{22} is the var of level 2 residuals u_{1i} from predicting the level quadrature slope (π_{2i})

$\text{cov}(u_{0i}, u_{1i}) = \text{cov}(\pi_{0i}, \pi_{1i})$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \\ e_{it} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00}^2 & \tau_{10}^2 & \tau_{20}^2 & 0 \\ \tau_{10}^2 & \tau_{11}^2 & \tau_{12}^2 & 0 \\ \tau_{20}^2 & \tau_{12}^2 & \tau_{22}^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}\right] \quad \text{(A2e)}$$

Without covariate terms:

$$\begin{pmatrix} \theta_{i0} \\ \theta_{i1} \\ : \\ : \\ \theta_{iT} \end{pmatrix} = \begin{pmatrix} 1 & \text{Year}_{i0} & \text{Year}_{i0}^2 \\ 1 & \text{Year}_{i1} & \text{Year}_{i1}^2 \\ : & : & : \\ : & : & : \\ 1 & \text{Year}_{iT} & \text{Year}_{iT}^2 \end{pmatrix} \begin{pmatrix} \beta_{00} \\ \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} 1 & \text{Year}_{i0} & \text{Year}_{i0}^2 \\ 1 & \text{Year}_{i1} & \text{Year}_{i1}^2 \\ : & : & : \\ : & : & : \\ 1 & \text{Year}_{iT} & \text{Year}_{iT}^2 \end{pmatrix} \begin{pmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{pmatrix} + \begin{pmatrix} e_{i0} \\ e_{i1} \\ : \\ : \\ e_{iT} \end{pmatrix} \quad \text{(A2f)}$$

$$\theta_i = \mathbf{X}_i\mathbf{\Gamma} + \mathbf{Z}_i\mathbf{U}_i + \mathbf{E}_i \quad \text{(A2g)}$$

where i =1,….N (number of person) and n = number of time points for individual i.
The marginal model (without Covariates) can be expressed:

$$\text{Var}(\theta_{it}) = \text{E}[(\theta_{it} - \text{E}(\theta_{it}))^2]$$
$$= \tau_0^2 + + \tau_1^2 \text{ Year}_{it}^2 + \tau_1^2 \text{ Year}_{it}^4 + 2\,\tau_{10} \text{ Year}_{it} + 2\,\tau_{20} \text{ Year}_{it}^2 + 2\,\tau_{12}\text{Year}_{it}^3 + \sigma^2 \quad \text{(A2h)}$$

$$\theta_{it} \sim N((\beta_{00} + \beta_{10} * \text{Year}_{it} + \beta_{20} * \text{Year}_{it}^2),\ \text{Var}(\theta_{it})) \quad \text{(A2i)}$$

C. Model for Serial Correlation

First order autoregressive process, AR1:

$E_{it} = \rho E_{i,(t-1)} + \varepsilon_{it}$

Where $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ and ρ is autocorrelation coefficient. In SAS,

$$\sigma^2 = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{ni-1} \\ \rho & 1 & \rho & \ldots & \rho^{ni-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{ni-3} \\ : & : & : & : & : \\ \rho^{ni-1} & \rho^{ni-2} & \rho^{ni-3} & \ldots & 1 \end{pmatrix}$$

Simulation Design

Table B1. Simulation Parameters (C ~ Cohort)

Grade	Linear theta	Non_Linear theta																			
1	-2	-2.00	C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0								
2	-1.6	-1.19		C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0							
3	-1.2	-0.53			C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0						
4	-0.8	0.07				C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0					
5	-0.4	0.59					C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0				
6	0	1.05						C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0			
7	0.4	1.44							C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0		
8	0.8	1.76								C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0	
9	1.2	2.00									C10	C9	C8	C7	C6	C5	C4	C3	C2	C1	C0
10	1.6	2.18										C10	C9	C8	C7	C6	C5	C4	C3	C2	C1
												Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
												Base									

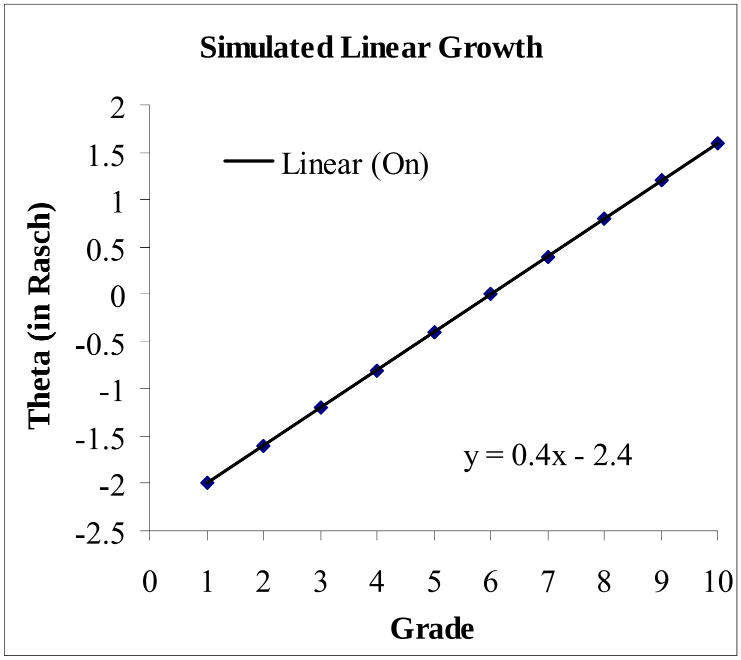


Figure B1. Simulated Linear Growth

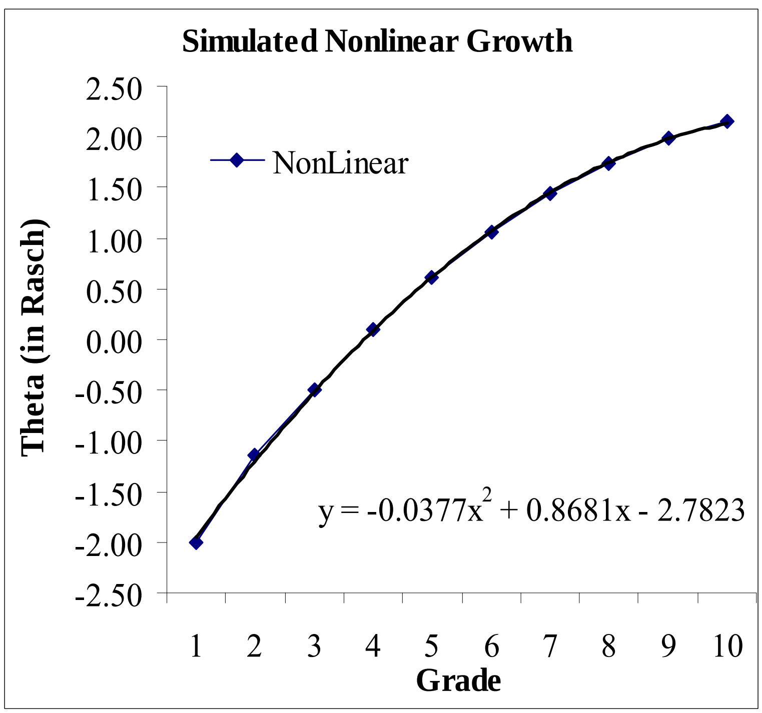


Figure B2. Simulated Nonlinear Growth

References

Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999).*Uncommon measures: Equivalence and linkage among educational tests*. Washington, DC: National Academy Press.

Camilli, G. (1999). Measurement error, multidimensionality, and scale shrinkage: A reply to Yen and Burket. Journal of Educational Measurement, 36, 73-78.

CTB/McGraw-Hill. (1997). *Winter norms book: TerraNova*. Monterey, CA: Author.

Diggle, P. J., Liang, K. L., & Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd edition. London: Oxford Science.

Doran, H. C. & Cohen, J. (2005). The confounding effect of linking bias on gains estimated from value-added models. In R. W. Lissitz (ed.), *Value-added models in educations: Theory and applications (pp. 80-104*). Maple Grove, MN: JAM Press.

Doran, H. C. & Jiang, T. (2006). The impact of linking error in longitudinal analysis: An empirical demonstration. In R. W. Lissitz (ed.), *Value-added models in educations*: *Theory and applications (pp. 21-229)*. Maple Grove, MN: JAM Press.

Goldstein, H. (1995). *Multilevel statistical models*, 2nd Edtion. London: Arnold.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications.* Boston: Kluwer.

Harcourt Educational Measurement. (2002). Stanford Achievement Test, 10th Edition. San Antonio, Texas.

Haertel, E. H. (1991). *Report on TRP analyses of issues concerning within-age versus cross-age scales for the National Assessment of Educational Progress*. Washington, D.C.: National Center for Educational Statistics.

Harwell, M. R., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement, 20,* 101–125.

Hoover, H. D., Dunbar, S. D., & Frisbie, D. A. (2003). *The Iowa Tests of Basic Skills. Interpretive guide for teachers and councelors. Forms A and B. Levels 9-14. Itasca, Il*: Riverside Publishing.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating: methods and practices* (2nd ed.). New York: Springer.

Linacre, J. M. (2008). WINSTEPS, program manual 3.67. Chicago IL.

Lissitz, R. W. & Huynh H. (2003). Vertical equating for state assessments: issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research & Evaluation, 8(10)*. Retrieved February 16, 2006 from http://PAREonline.net/getvn.asp?v=8&n=10.

Little, R.C., Milliken, G.A.., Stroup, W.W., & Wolfinger, R.D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.

Lockwood, J. & McCaffrey, D. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics, vol. 1*. 223–252

Martineau, J. (2006). Distorting value-added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Psychological Statistics*, pp 35-62.

Molenburgs, G. & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. NY: Springer.

Robin, F., Holland P, & Hemat, L. (2006). ICEDOG software.

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

SAS Institute Inc. (2002-2003). SAS (version 9). Cary, NC, USA.

Schafer, W. D. (2006). Growth Scales as an Alternative to Vertical Scales. *Practical Assessment Research & Evaluation, 11(4)*. Available online: http://pareonline.net/getvn.asp?v=11&

Stanford Achievement Test Series, *Tenth Edition* (Technical Data Report). (2004). San Antonio, TX: Harcourt Assessment, Inc.

Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. Applied Psychological Measurement, 7, 201-210.

Yovanoff, P., Duesbery, L., Alonzo, J. & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and Oral reading fluency. *Educational Measurement: Issues and Practice*. *24* (3) 4-10.

Table 1. Example of Vertical Scale across Grades for Different Cohort Groups (C_1 to C_{10})

					Year				
Grade	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
1	$\mathbf{C_1}$	$\mathbf{C_0}$	$\mathbf{C_{-1}}$	$\mathbf{C_{-2}}$	$\mathbf{C_{-3}}$					
2	$\mathbf{C_2}$	$\mathbf{C_1}$	$\mathbf{C_0}$	$\mathbf{C_{-1}}$	$\mathbf{C_{-2}}$	C_{-3}				
3	$\mathbf{C_3}$	$\mathbf{C_2}$	$\mathbf{C_1}$	$\mathbf{C_0}$	$\mathbf{C_{-1}}$	C_{-2}	C_{-3}			
4	$\mathbf{C_4}$	$\mathbf{C_3}$	$\mathbf{C_2}$	$\mathbf{C_1}$	$\mathbf{C_0}$	C_{-1}	C_{-2}	C_{-3}		
5	$\mathbf{C_5}$	$\mathbf{C_4}$	$\mathbf{C_3}$	$\mathbf{C_2}$	$\mathbf{C_1}$	C_0	C_{-1}	C_{-2}	C_{-3}	
6	$\mathbf{C_6}$	$\mathbf{C_5}$	$\mathbf{C_4}$	$\mathbf{C_3}$	$\mathbf{C_2}$	C_1	C_0	C_{-1}	C_{-2}	C_{-3}
7	$\mathbf{C_7}$	$\mathbf{C_6}$	$\mathbf{C_5}$	$\mathbf{C_4}$	$\mathbf{C_3}$	C_2	C_1	C_0	C_{-1}	C_{-2}
8	$\mathbf{C_8}$	$\mathbf{C_7}$	$\mathbf{C_6}$	$\mathbf{C_5}$	$\mathbf{C_4}$	C_3	C_2	C_1	C_0	C_{-1}
9	$\mathbf{C_9}$	$\mathbf{C_8}$	$\mathbf{C_7}$	$\mathbf{C_6}$	$\mathbf{C_5}$	C_4	C_3	C_2	C_1	C_0
10	$\mathbf{C_{10}}$	$\mathbf{C_9}$	$\mathbf{C_8}$	$\mathbf{C_7}$	$\mathbf{C_6}$	C_5	C_4	C_3	C_2	C_1

Table 2a. Vertical Scaling Linking Design for On-grade and Below-grade Items for Base Year (Year 1).

	Item									
Grade	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	G1_on									
G2	G1_below	G2_on								
G3		G2_below	G3_on							
G4			G3_below	G4_on						
G5				G4_below	G5_on					
G6					G5_below	G6_on				
G7						G6_below	G7_on			
G8							G7_below	G8_on		
G9								G8_below	G9_on	
G10									G9_below	G10_on

Table 2b. Examples of Common-Item Chain Equating Design for On-grade Items of Different Years Linked to Base Year for Grade 1

Grade	Year 1	Year 1 & Year 2	Year 2
G1	30 items	10 items	
		10 items	30 items
	Year 2	Year 2 & Year 3	Year 3
G1	30 items	10 items	
		10 items	30 items
	Year 3	Year 3 & Year 4	Year 4
G1	30 items	10 items	
		10 items	30 items
	Year 4	Year 4 & Year 5	Year 5
G1	30 items	10 items	
		10 items	30 items

Table 3. Simulation Rasch difficulty Parameters of 40 Item Test Forms across Grades ~ *N*(M, SD)

Grade	Off_mean	On_mean
1		-2
2	-2	-1.6
3	-1.6	-1.2
4	-1.2	-0.8
5	-0.8	-0.4
6	-0.4	0
7	0	0.4
8	0.4	0.8
9	0.8	1.2
10	1.2	1.6

Table 4. Bias, SE, and RMSE for Linear Null, Intercept, and Intercept + Slope Models Without Noise

Linear		Null					Intercept					Intercept + Slope				
	Grade	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5
Bias	1	0.02	0.03	0.07	0.04	0.06	0.03	0.00	0.01	0.01	0.01	0.02	0.04	0.02	0.02	-0.01
	2	0.03	0.00	0.03	0.07	0.03	0.04	0.05	0.02	0.00	0.01	0.05	0.01	0.04	0.04	0.04
	3	0.03	0.03	0.00	0.03	0.04	0.07	0.06	0.05	0.02	0.00	0.03	-0.01	0.04	0.09	0.02
	4	0.04	0.08	0.06	0.05	0.10	0.09	0.11	0.09	0.10	0.06	0.10	0.08	0.05	0.16	0.09
	5	0.05	0.02	0.05	0.06	0.04	0.06	0.05	0.09	0.06	0.05	0.10	0.06	0.05	0.04	0.08
	6	0.06	0.06	0.01	0.03	0.07	0.05	0.08	0.05	0.06	0.10	0.10	0.05	0.08	0.04	0.00
	7	0.06	0.08	0.09	0.04	0.07	0.09	0.04	0.14	0.09	0.08	0.10	0.10	0.09	0.17	0.03
	8	0.07	0.03	0.04	0.06	0.05	0.06	0.05	-0.01	0.07	0.04	0.02	0.07	0.08	0.06	0.10
	9	0.07	0.11	0.11	0.13	0.08	0.05	0.11	0.13	0.05	0.08	0.11	0.07	0.10	0.08	0.13
	10	0.08	0.02	0.11	0.09	0.13	0.02	0.00	0.09	0.12	0.02	0.09	0.05	0.02	0.00	0.05
	Mean	0.05	0.05	0.06	0.06	0.07	0.06	0.05	0.07	0.06	0.05	0.07	0.05	0.06	0.07	0.05
SE	1	0.00	0.04	0.05	0.06	0.07	0.01	0.03	0.03	0.04	0.05	0.01	0.02	0.03	0.04	0.05
	2	0.02	0.03	0.04	0.05	0.05	0.02	0.03	0.04	0.05	0.05	0.02	0.03	0.05	0.05	0.06
	3	0.03	0.04	0.05	0.05	0.05	0.03	0.03	0.04	0.05	0.05	0.02	0.04	0.05	0.06	0.06
	4	0.03	0.04	0.05	0.05	0.05	0.03	0.03	0.05	0.06	0.06	0.03	0.03	0.04	0.05	0.06
	5	0.03	0.04	0.05	0.05	0.05	0.03	0.05	0.05	0.06	0.06	0.03	0.04	0.04	0.05	0.07
	6	0.03	0.03	0.05	0.05	0.06	0.03	0.04	0.05	0.05	0.05	0.03	0.04	0.06	0.06	0.06
	7	0.03	0.04	0.05	0.06	0.06	0.03	0.04	0.05	0.05	0.06	0.03	0.04	0.06	0.05	0.06
	8	0.04	0.05	0.05	0.05	0.06	0.04	0.05	0.06	0.06	0.06	0.03	0.04	0.05	0.05	0.06
	9	0.04	0.04	0.05	0.05	0.06	0.04	0.04	0.05	0.06	0.07	0.04	0.05	0.06	0.07	0.08
	10	0.04	0.05	0.05	0.06	0.07	0.04	0.04	0.05	0.06	0.06	0.04	0.05	0.06	0.09	0.09
	Mean	0.03	0.04	0.05	0.05	0.06	0.03	0.04	0.05	0.05	0.06	0.03	0.04	0.05	0.06	0.07
RMSE	1	0.02	0.05	0.08	0.07	0.09	0.03	0.03	0.04	0.04	0.05	0.02	0.04	0.04	0.04	0.05
	2	0.03	0.03	0.05	0.08	0.06	0.04	0.06	0.04	0.05	0.06	0.05	0.03	0.06	0.07	0.07
	3	0.04	0.05	0.05	0.06	0.07	0.08	0.07	0.07	0.06	0.05	0.04	0.04	0.06	0.11	0.06
	4	0.05	0.09	0.07	0.07	0.11	0.09	0.11	0.10	0.12	0.08	0.10	0.09	0.06	0.16	0.11
	5	0.06	0.04	0.07	0.08	0.07	0.07	0.07	0.10	0.08	0.08	0.11	0.07	0.07	0.07	0.11
	6	0.06	0.07	0.05	0.06	0.09	0.06	0.09	0.07	0.08	0.12	0.10	0.07	0.10	0.07	0.06
	7	0.07	0.09	0.10	0.07	0.10	0.09	0.06	0.15	0.10	0.10	0.11	0.11	0.10	0.18	0.07
	8	0.08	0.05	0.07	0.08	0.08	0.07	0.07	0.06	0.09	0.07	0.04	0.08	0.09	0.08	0.11
	9	0.08	0.12	0.12	0.14	0.10	0.06	0.12	0.13	0.08	0.11	0.12	0.08	0.12	0.10	0.15
	10	0.09	0.05	0.12	0.11	0.14	0.05	0.04	0.10	0.14	0.06	0.09	0.07	0.07	0.09	0.10
	Mean	0.06	0.06	0.08	0.08	0.09	0.06	0.07	0.09	0.08	0.08	0.08	0.07	0.08	0.10	0.09

Table 5. Bias, SE, and RMSE for Linear Null, Intercept, and Intercept + Slope Models With Noise*

Linear	Null Model						Intercept Model					Intercept + Slope Model				
	Grade	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5
Bias	1	0.02	0.03	0.07	0.04	0.06	0.04	-0.10	-0.09	-0.08	-0.08	0.03	-0.06	-0.06	-0.07	-0.10
	2	0.03	0.00	0.03	0.07	0.03	-0.06	0.06	-0.07	-0.09	-0.07	-0.05	0.07	-0.05	-0.05	-0.06
	3	0.03	0.03	0.00	0.03	0.04	-0.03	-0.03	0.07	-0.06	-0.09	-0.06	-0.11	0.14	-0.01	-0.10
	4	0.04	0.08	0.06	0.05	0.10	0.00	0.01	0.00	0.12	-0.04	0.00	-0.03	-0.05	0.27	-0.01
	5	0.05	0.02	0.05	0.06	0.04	-0.02	-0.04	-0.01	-0.04	0.06	0.00	-0.03	-0.04	-0.05	0.22
	6	0.06	0.06	0.01	0.03	0.07	-0.14	-0.01	-0.04	-0.03	0.01	-0.10	-0.06	-0.03	-0.07	-0.12
	7	0.06	0.08	0.09	0.04	0.07	-0.01	-0.15	0.05	0.00	-0.02	0.00	-0.16	0.00	0.08	-0.05
	8	0.07	0.03	0.04	0.06	0.05	-0.03	-0.03	-0.20	-0.02	-0.04	-0.09	-0.03	-0.22	-0.05	-0.01
	9	0.07	0.11	0.11	0.13	0.08	-0.03	0.03	0.04	-0.14	0.00	0.01	-0.04	0.00	-0.26	0.02
	10	0.08	0.02	0.11	0.09	0.13	-0.06	-0.09	0.00	0.04	-0.17	-0.01	-0.05	-0.07	-0.11	-0.31
	Mean	**0.05**	**0.05**	**0.06**	**0.06**	**0.07**	**-0.03**	**-0.04**	**-0.02**	**-0.03**	**-0.04**	**-0.03**	**-0.05**	**-0.04**	**-0.03**	**-0.05**
SE	1	0.00	0.04	0.05	0.06	0.07	0.01	0.02	0.04	0.04	0.06	0.01	0.02	0.04	0.04	0.05
	2	0.02	0.03	0.04	0.05	0.05	0.01	0.03	0.04	0.04	0.11	0.02	0.03	0.04	0.05	0.05
	3	0.03	0.04	0.05	0.05	0.05	0.02	0.04	0.05	0.05	0.06	0.03	0.03	0.05	0.06	0.07
	4	0.03	0.04	0.05	0.05	0.05	0.02	0.03	0.04	0.04	0.05	0.03	0.04	0.05	0.07	0.07
	5	0.03	0.04	0.05	0.05	0.05	0.03	0.03	0.04	0.04	0.05	0.03	0.06	0.07	0.08	0.08
	6	0.03	0.03	0.05	0.05	0.06	0.03	0.04	0.04	0.06	0.06	0.03	0.04	0.05	0.05	0.07
	7	0.03	0.04	0.05	0.06	0.06	0.03	0.04	0.04	0.05	0.06	0.04	0.04	0.05	0.06	0.07
	8	0.04	0.05	0.05	0.05	0.06	0.04	0.05	0.05	0.06	0.06	0.04	0.05	0.06	0.07	0.07
	9	0.04	0.04	0.05	0.05	0.06	0.04	0.05	0.05	0.06	0.06	0.04	0.05	0.06	0.07	0.07
	10	0.04	0.05	0.05	0.06	0.07	0.04	0.05	0.05	0.06	0.07	0.04	0.05	0.06	0.06	0.07
	Mean	**0.03**	**0.04**	**0.05**	**0.05**	**0.06**	**0.03**	**0.04**	**0.04**	**0.05**	**0.06**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**
RMSE	1	0.02	0.05	0.08	0.07	0.09	0.04	0.10	0.09	0.09	0.10	0.03	0.06	0.07	0.08	0.11
	2	0.03	0.03	0.05	0.08	0.06	0.06	0.07	0.08	0.10	0.13	0.05	0.08	0.07	0.07	0.08
	3	0.04	0.05	0.05	0.06	0.07	0.03	0.05	0.08	0.08	0.11	0.07	0.12	0.14	0.06	0.12
	4	0.05	0.09	0.07	0.07	0.11	0.02	0.04	0.04	0.12	0.07	0.03	0.05	0.07	0.28	0.07
	5	0.06	0.04	0.07	0.08	0.07	0.04	0.05	0.04	0.06	0.08	0.03	0.06	0.08	0.09	0.23
	6	0.06	0.07	0.05	0.06	0.09	0.15	0.04	0.06	0.06	0.06	0.11	0.07	0.06	0.09	0.14
	7	0.07	0.09	0.10	0.07	0.10	0.03	0.15	0.06	0.05	0.06	0.04	0.16	0.05	0.10	0.09
	8	0.08	0.05	0.07	0.08	0.08	0.05	0.06	0.20	0.06	0.07	0.10	0.06	0.23	0.08	0.07
	9	0.08	0.12	0.12	0.14	0.10	0.05	0.06	0.07	0.15	0.06	0.04	0.06	0.06	0.26	0.08
	10	0.09	0.05	0.12	0.11	0.14	0.07	0.10	0.05	0.07	0.18	0.05	0.07	0.10	0.12	0.32
	Mean	**0.06**	**0.06**	**0.08**	**0.08**	**0.09**	**0.05**	**0.07**	**0.08**	**0.08**	**0.09**	**0.05**	**0.08**	**0.09**	**0.12**	**0.13**

*: There is no noise in the null model

Table 6. Bias, SE, and RMSE for Nonlinear Null, Intercept, and Intercept + Slope Models without Noise

Nonlinear		Null					Intercept					Intercept + Slope				
	Grade	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5
Bias	1	-0.02	-0.05	-0.05	-0.04	-0.01	-0.01	0.01	-0.02	-0.07	-0.09	-0.01	0.02	-0.03	-0.02	0.03
	2	0.01	0.02	-0.04	-0.02	-0.01	-0.09	-0.01	0.03	0.01	-0.07	-0.04	0.09	0.09	0.01	-0.01
	3	0.02	0.06	0.06	0.01	0.01	0.06	-0.08	0.02	0.06	0.02	0.08	-0.02	0.13	0.07	-0.07
	4	0.03	0.06	0.08	0.07	0.04	-0.01	0.03	-0.06	0.01	0.08	0.02	0.14	-0.09	-0.05	0.06
	5	0.04	0.06	0.07	0.09	0.10	0.09	-0.01	0.02	0.02	0.08	0.02	0.06	0.14	-0.28	-0.22
	6	0.04	0.05	0.08	0.08	0.11	0.13	0.09	0.03	0.06	0.08	0.08	0.02	-0.01	-0.05	-0.35
	7	0.05	0.06	0.07	0.07	0.07	0.04	0.13	0.07	0.05	0.01	0.08	0.09	-0.05	-0.23	-0.22
	8	0.05	0.06	0.04	0.05	0.05	0.03	0.07	0.13	0.07	0.03	0.02	0.08	0.01	-0.23	-0.39
	9	0.05	0.11	0.09	0.09	0.10	0.06	0.13	0.13	0.14	0.14	0.07	0.05	0.00	-0.08	-0.19
	10	0.05	0.02	0.07	0.06	0.02	0.00	-0.01	0.06	0.06	0.05	0.02	0.02	-0.03	-0.15	-0.20
	Mean	**0.03**	**0.05**	**0.05**	**0.05**	**0.05**	**0.03**	**0.04**	**0.04**	**0.04**	**0.03**	**0.03**	**0.06**	**0.02**	**-0.10**	**-0.16**
SE	1	0.00	0.04	0.04	0.05	0.06	0.01	0.03	0.04	0.04	0.04	0.01	0.02	0.03	0.04	0.04
	2	0.02	0.04	0.04	0.05	0.06	0.02	0.04	0.05	0.05	0.06	0.02	0.03	0.05	0.06	0.07
	3	0.02	0.03	0.04	0.05	0.05	0.02	0.03	0.04	0.07	0.07	0.02	0.03	0.04	0.06	0.06
	4	0.02	0.03	0.05	0.05	0.06	0.03	0.04	0.04	0.05	0.06	0.02	0.04	0.05	0.06	0.07
	5	0.03	0.04	0.05	0.06	0.06	0.03	0.04	0.05	0.05	0.06	0.03	0.04	0.05	0.06	0.07
	6	0.03	0.04	0.04	0.06	0.06	0.04	0.04	0.05	0.07	0.08	0.04	0.04	0.06	0.06	0.06
	7	0.03	0.04	0.04	0.05	0.05	0.04	0.04	0.06	0.06	0.06	0.04	0.05	0.06	0.07	0.07
	8	0.03	0.04	0.04	0.05	0.05	0.04	0.06	0.07	0.07	0.08	0.04	0.05	0.06	0.08	0.09
	9	0.04	0.04	0.05	0.06	0.07	0.04	0.05	0.06	0.06	0.07	0.04	0.04	0.05	0.06	0.08
	10	0.04	0.04	0.05	0.05	0.06	0.04	0.05	0.05	0.06	0.07	0.04	0.05	0.05	0.07	0.09
	Mean	**0.03**	**0.04**	**0.04**	**0.05**	**0.06**	**0.03**	**0.04**	**0.05**	**0.06**	**0.06**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**
RMSE	1	0.02	0.06	0.07	0.06	0.06	0.01	0.03	0.04	0.08	0.10	0.02	0.03	0.04	0.04	0.05
	2	0.02	0.04	0.06	0.05	0.06	0.09	0.04	0.06	0.05	0.09	0.05	0.09	0.10	0.06	0.07
	3	0.03	0.07	0.07	0.05	0.05	0.07	0.08	0.05	0.09	0.08	0.08	0.03	0.14	0.09	0.09
	4	0.04	0.07	0.09	0.09	0.07	0.03	0.05	0.08	0.05	0.10	0.03	0.14	0.10	0.08	0.10
	5	0.05	0.07	0.09	0.11	0.12	0.09	0.04	0.05	0.05	0.10	0.03	0.08	0.15	0.29	0.23
	6	0.05	0.06	0.09	0.10	0.12	0.14	0.10	0.06	0.10	0.11	0.09	0.05	0.06	0.08	0.36
	7	0.06	0.07	0.09	0.09	0.08	0.05	0.13	0.09	0.08	0.06	0.09	0.10	0.08	0.24	0.23
	8	0.06	0.07	0.06	0.07	0.07	0.05	0.09	0.15	0.10	0.08	0.05	0.09	0.06	0.24	0.40
	9	0.06	0.12	0.10	0.11	0.12	0.07	0.14	0.14	0.15	0.15	0.08	0.07	0.05	0.10	0.20
	10	0.07	0.04	0.08	0.08	0.07	0.04	0.05	0.08	0.08	0.08	0.04	0.05	0.06	0.17	0.22
	Mean	**0.04**	**0.07**	**0.08**	**0.08**	**0.08**	**0.07**	**0.08**	**0.08**	**0.08**	**0.10**	**0.06**	**0.07**	**0.08**	**0.14**	**0.20**

Table 7. Bias, SE, and RMSE for Nonlinear Null, Intercept, and Intercept + Slope Models with Noise*

Nonlinear		Null					Intercept					Intercept + Slope				
	Grade	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5
Bias	1	-0.02	-0.05	-0.05	-0.04	-0.01	-0.01	-0.09	-0.12	-0.17	-0.19	-0.01	-0.06	-0.10	-0.10	-0.05
	2	0.01	0.02	-0.04	-0.02	-0.01	-0.19	-0.01	-0.07	-0.09	-0.16	-0.13	0.15	0.01	-0.09	-0.12
	3	0.02	0.06	0.06	0.01	0.01	-0.04	-0.18	0.02	-0.05	-0.09	-0.02	-0.10	0.25	-0.02	-0.16
	4	0.03	0.06	0.08	0.07	0.04	-0.11	-0.07	-0.16	0.01	-0.02	-0.07	0.05	-0.16	0.07	-0.01
	5	0.04	0.06	0.07	0.09	0.10	-0.01	-0.11	-0.07	-0.09	0.08	-0.08	-0.03	0.06	-0.37	-0.13
	6	0.04	0.05	0.08	0.08	0.11	-0.07	-0.01	-0.07	-0.05	-0.03	-0.11	-0.07	-0.10	-0.13	-0.45
	7	0.05	0.06	0.07	0.07	0.07	-0.07	-0.08	-0.03	-0.05	-0.10	-0.01	-0.16	-0.13	-0.31	-0.29
	8	0.05	0.06	0.04	0.05	0.05	-0.07	-0.04	-0.09	-0.04	-0.09	-0.07	-0.03	-0.28	-0.32	-0.49
	9	0.05	0.11	0.09	0.09	0.10	-0.04	0.02	0.02	-0.09	0.01	-0.03	-0.04	-0.08	-0.36	-0.26
	10	0.05	0.02	0.07	0.06	0.02	-0.11	-0.11	-0.04	-0.05	-0.16	-0.08	-0.06	-0.11	-0.22	-0.46
	Mean	**0.03**	**0.05**	**0.05**	**0.05**	**0.05**	**-0.07**	**-0.07**	**-0.06**	**-0.07**	**-0.07**	**-0.06**	**-0.04**	**-0.07**	**-0.19**	**-0.24**
SE	1	0.00	0.04	0.04	0.05	0.06	0.01	0.02	0.03	0.04	0.05	0.01	0.04	0.05	0.06	0.06
	2	0.02	0.04	0.04	0.05	0.06	0.01	0.04	0.04	0.05	0.06	0.02	0.03	0.04	0.04	0.05
	3	0.02	0.03	0.04	0.05	0.05	0.02	0.04	0.04	0.05	0.05	0.02	0.03	0.04	0.05	0.06
	4	0.02	0.03	0.05	0.05	0.06	0.03	0.03	0.04	0.05	0.06	0.03	0.04	0.05	0.07	0.08
	5	0.03	0.04	0.05	0.06	0.06	0.03	0.04	0.05	0.05	0.06	0.03	0.04	0.05	0.07	0.07
	6	0.03	0.04	0.04	0.06	0.06	0.04	0.05	0.05	0.06	0.07	0.04	0.04	0.05	0.06	0.07
	7	0.03	0.04	0.04	0.05	0.05	0.04	0.05	0.06	0.07	0.08	0.04	0.05	0.06	0.08	0.07
	8	0.03	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.07	0.09	0.04	0.05	0.06	0.06	0.08
	9	0.04	0.04	0.05	0.06	0.07	0.05	0.06	0.07	0.08	0.08	0.04	0.05	0.07	0.07	0.07
	10	0.04	0.04	0.05	0.05	0.06	0.06	0.06	0.07	0.08	0.09	0.04	0.05	0.06	0.06	0.07
	Mean	**0.03**	**0.04**	**0.04**	**0.05**	**0.06**	**0.03**	**0.05**	**0.05**	**0.06**	**0.07**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**
RMSE	1	0.02	0.06	0.07	0.06	0.06	0.01	0.10	0.13	0.17	0.19	0.01	0.07	0.12	0.11	0.08
	2	0.02	0.04	0.06	0.05	0.06	0.19	0.04	0.08	0.11	0.17	0.14	0.16	0.04	0.10	0.13
	3	0.03	0.07	0.07	0.05	0.05	0.05	0.19	0.05	0.06	0.10	0.03	0.11	0.25	0.06	0.18
	4	0.04	0.07	0.09	0.09	0.07	0.11	0.08	0.17	0.05	0.06	0.08	0.06	0.17	0.10	0.08
	5	0.05	0.07	0.09	0.11	0.12	0.03	0.12	0.08	0.10	0.10	0.08	0.05	0.08	0.38	0.15
	6	0.05	0.06	0.09	0.10	0.12	0.08	0.05	0.09	0.07	0.07	0.12	0.08	0.11	0.14	0.46
	7	0.06	0.07	0.09	0.09	0.08	0.08	0.09	0.07	0.09	0.13	0.04	0.17	0.15	0.32	0.30
	8	0.06	0.07	0.06	0.07	0.07	0.09	0.07	0.11	0.08	0.12	0.08	0.05	0.29	0.33	0.50
	9	0.06	0.12	0.10	0.11	0.12	0.07	0.07	0.07	0.12	0.08	0.05	0.07	0.11	0.37	0.27
	10	0.07	0.04	0.08	0.08	0.07	0.12	0.13	0.08	0.09	0.18	0.09	0.08	0.13	0.23	0.47
	Mean	**0.04**	**0.07**	**0.08**	**0.08**	**0.08**	0.08	**0.09**	**0.09**	**0.10**	**0.12**	**0.07**	**0.09**	**0.14**	**0.21**	**0.26**

*: There is no noise in the null model

Table 8. Vertical Scaling Constants of Linear and Nonlinear Null, Intercept, and Intercept + Slope Models without and with Noise

		Null					Intercept					Intercept + Slope				
	Grade	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5	Year1	Year2	Year3	Year4	Year5
Linear	1	-2.00	-2.00	-2.00	-2.00	-2.00	-1.99	-2.02	-2.01	-2.00	-2.01	-1.99	-1.99	-1.99	-2.00	-2.02
without	2	-1.60	-1.60	-1.60	-1.60	-1.60	-1.58	-1.62	-1.60	-1.62	-1.62	-1.57	-1.61	-1.58	-1.57	-1.59
Noise	3	-1.20	-1.20	-1.20	-1.20	-1.20	-1.14	-1.20	-1.16	-1.22	-1.25	-1.18	-1.24	-1.24	-1.11	-1.21
	4	-0.80	-0.80	-0.80	-0.80	-0.80	-0.76	-0.80	-0.79	-0.81	-0.85	-0.75	-0.83	-0.87	-0.72	-0.77
	5	-0.40	-0.40	-0.40	-0.40	-0.40	-0.34	-0.40	-0.36	-0.42	-0.40	-0.31	-0.39	-0.43	-0.37	-0.26
	6	0.00	0.00	0.00	0.00	0.00	0.06	0.04	0.02	0.02	0.05	0.11	-0.02	-0.02	0.07	-0.06
	7	0.40	0.40	0.40	0.40	0.40	0.47	0.38	0.48	0.42	0.40	0.49	0.41	0.37	0.60	0.32
	8	0.80	0.80	0.80	0.80	0.80	0.87	0.79	0.76	0.85	0.82	0.83	0.80	0.82	0.91	0.82
	9	1.20	1.20	1.20	1.20	1.20	1.22	1.17	1.26	1.17	1.22	1.28	1.15	1.16	1.32	1.23
	10	1.60	1.60	1.60	1.60	1.60	1.63	1.50	1.65	1.67	1.55	1.69	1.61	1.52	1.58	1.72
Nonlinear	1	-2.00	-2.00	-2.00	-2.00	-2.00	-1.99	-1.97	-1.99	-2.05	-2.07	-1.99	-1.96	-2.00	-2.00	-1.94
without	2	-1.19	-1.19	-1.19	-1.19	-1.19	-1.27	-1.19	-1.14	-1.17	-1.26	-1.22	-1.08	-1.10	-1.16	-1.19
Noise	3	-0.53	-0.53	-0.53	-0.53	-0.53	-0.46	-0.62	-0.49	-0.48	-0.52	-0.45	-0.53	-0.41	-0.42	-0.61
	4	0.07	0.07	0.07	0.07	0.07	0.05	0.06	0.01	0.04	0.13	0.07	0.23	0.01	0.08	0.08
	5	0.59	0.59	0.59	0.59	0.59	0.67	0.54	0.64	0.57	0.66	0.60	0.68	0.71	0.30	0.52
	6	1.05	1.05	1.05	1.05	1.05	1.16	1.07	1.06	1.04	1.10	1.11	1.11	0.97	0.96	0.67
	7	1.44	1.44	1.44	1.44	1.44	1.46	1.51	1.51	1.46	1.44	1.50	1.55	1.33	1.24	1.23
	8	1.76	1.76	1.76	1.76	1.76	1.79	1.78	1.92	1.80	1.78	1.79	1.88	1.74	1.52	1.36
	9	2.00	2.00	2.00	2.00	2.00	2.05	2.05	2.17	2.08	2.11	2.06	2.05	1.96	1.99	1.74
	10	2.18	2.18	2.18	2.18	2.18	2.21	2.14	2.34	2.26	2.25	2.23	2.24	2.19	2.11	2.07
Linear	1	-2.00	-2.00	-2.00	-2.00	-2.00	-1.98	-2.12	-2.11	-2.10	-2.11	-1.99	-2.09	-2.09	-2.09	-2.12
with	2	-1.60	-1.60	-1.60	-1.60	-1.60	-1.67	-1.55	-1.67	-1.69	-1.68	-1.67	-1.56	-1.66	-1.66	-1.67
Noise	3	-1.20	-1.20	-1.20	-1.20	-1.20	-1.24	-1.27	-1.11	-1.27	-1.30	-1.27	-1.34	-1.11	-1.21	-1.33
	4	-0.80	-0.80	-0.80	-0.80	-0.80	-0.85	-0.90	-0.81	-0.75	-0.88	-0.85	-0.87	-0.92	-0.52	-0.91
	5	-0.40	-0.40	-0.40	-0.40	-0.40	-0.43	-0.52	-0.37	-0.47	-0.37	-0.41	-0.46	-0.47	-0.47	-0.23
	6	0.00	0.00	0.00	0.00	0.00	-0.14	-0.07	0.00	-0.05	-0.02	-0.10	0.01	-0.03	-0.02	-0.20
	7	0.40	0.40	0.40	0.40	0.40	0.38	0.19	0.45	0.37	0.34	0.39	0.28	0.36	0.50	0.21
	8	0.80	0.80	0.80	0.80	0.80	0.78	0.74	0.63	0.77	0.74	0.72	0.83	0.59	0.80	0.73
	9	1.20	1.20	1.20	1.20	1.20	1.14	1.16	1.23	0.99	1.15	1.18	1.18	1.20	0.97	1.17
	10	2.18	2.18	2.18	2.18	2.18	1.55	1.48	1.65	1.62	1.39	1.59	1.61	1.60	1.60	1.22
Nonlinear	1	-2.00	-2.00	-2.00	-2.00	-2.00	-1.99	-2.07	-2.09	-2.15	-2.17	-1.99	-2.05	-2.09	-2.09	-2.04
with	2	-1.19	-1.19	-1.19	-1.19	-1.19	-1.36	-1.18	-1.25	-1.27	-1.33	-1.31	-1.02	-1.16	-1.26	-1.30
Noise	3	-0.53	-0.53	-0.53	-0.53	-0.53	-0.57	-0.71	-0.49	-0.57	-0.63	-0.54	-0.65	-0.24	-0.54	-0.70
	4	0.07	0.07	0.07	0.07	0.07	-0.05	0.00	-0.11	0.05	0.02	-0.02	0.09	-0.01	0.18	0.03
	5	0.59	0.59	0.59	0.59	0.59	0.57	0.50	0.48	0.45	0.61	0.51	0.52	0.66	0.19	0.57
	6	1.05	1.05	1.05	1.05	1.05	0.96	1.06	0.96	0.92	0.96	0.92	0.95	0.96	0.92	0.52
	7	1.44	1.44	1.44	1.44	1.44	1.36	1.40	1.40	1.32	1.31	1.41	1.24	1.32	1.12	1.09
	8	1.76	1.76	1.76	1.76	1.76	1.69	1.78	1.71	1.66	1.63	1.69	1.71	1.55	1.40	1.38
	9	2.00	2.00	2.00	2.00	2.00	1.94	2.02	2.03	1.83	1.98	1.96	1.92	1.92	1.69	1.68
	10	2.18	2.18	2.18	2.18	2.18	2.10	2.10	2.20	2.09	2.03	2.13	2.12	2.16	1.99	1.86

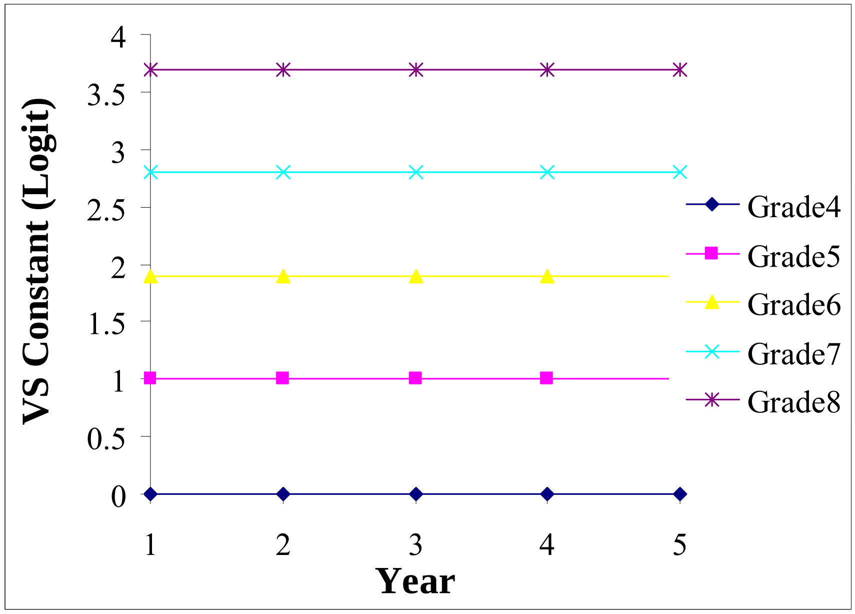


Figure 1. Static Vertical Scale across Years

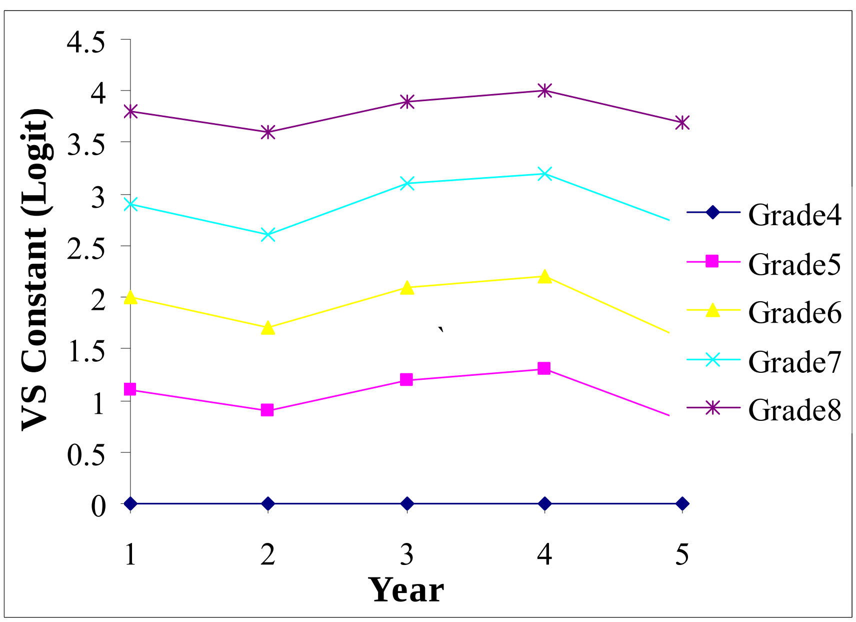


Figure 2. Dynamic Vertical Scale across Year

Linear Null Model

Linear Intercept (Without Noise)

Linear Intercept + Slope (Without Noise)

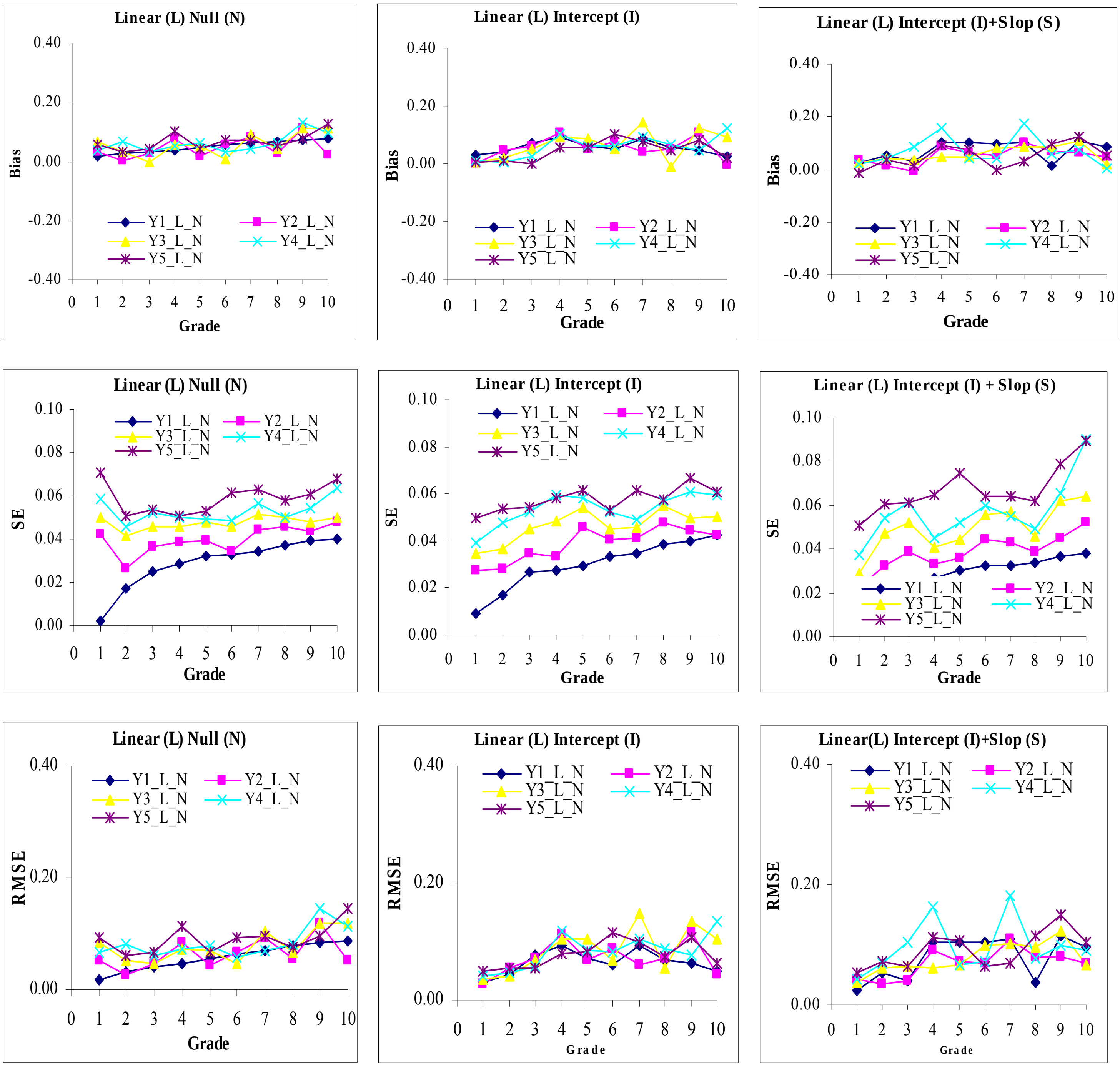


Figure 3. Bias, SE, and RMSE of Linear Models (Null, Intercept, and Intercept + Slope) without Noise

Linear Null Model

Linear Intercept (With Noise)

Linear Intercept + Slope (With Noise)

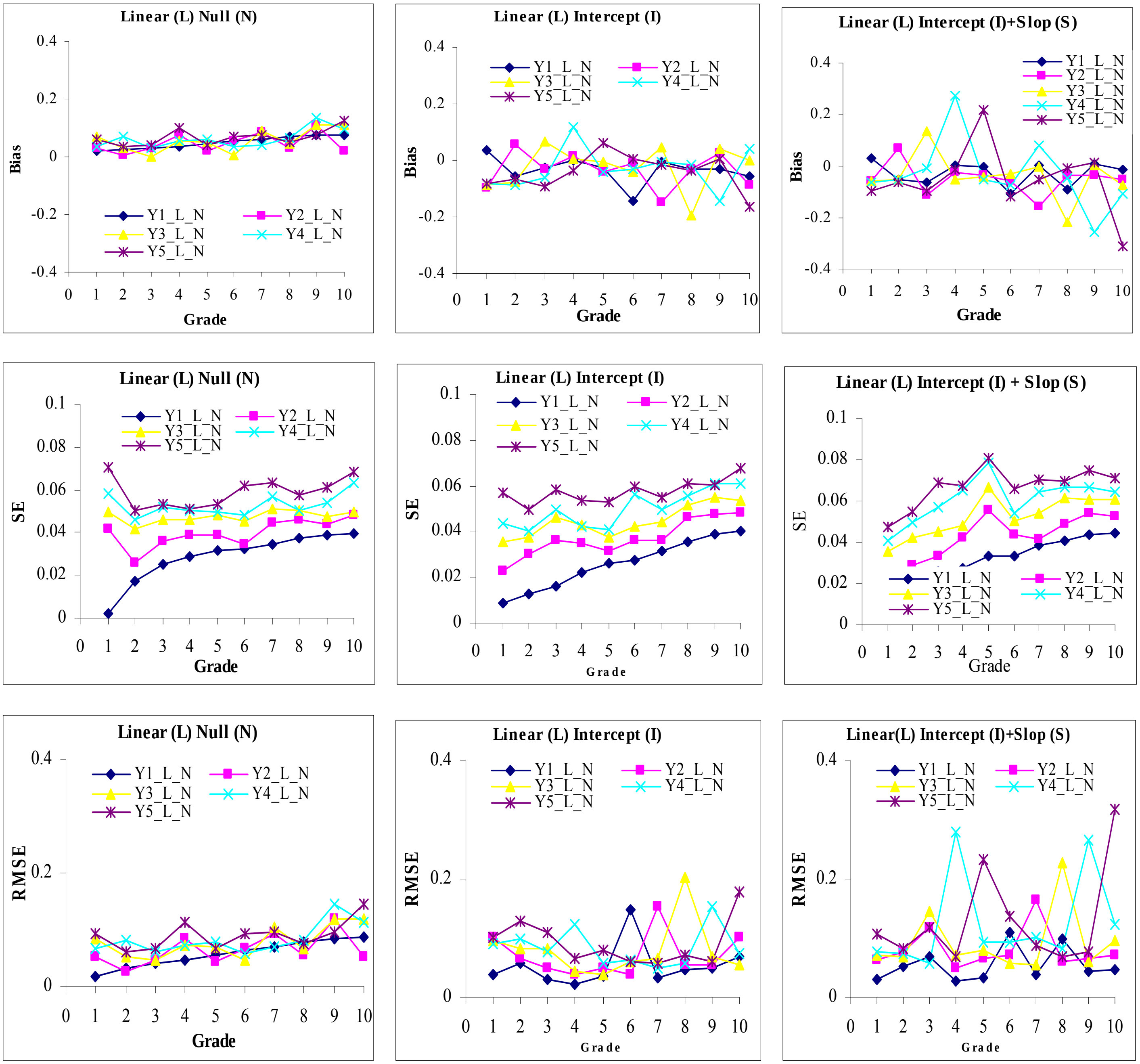


Figure 4. Bias, SE, and RMSE of Linear Models (Null, Intercept, and Intercept + Slope) with Noise

Nonlinear Null Model

Nonlinear Intercept (Without Noise)

Nonlinear Intercept + Slope (Without Noise)

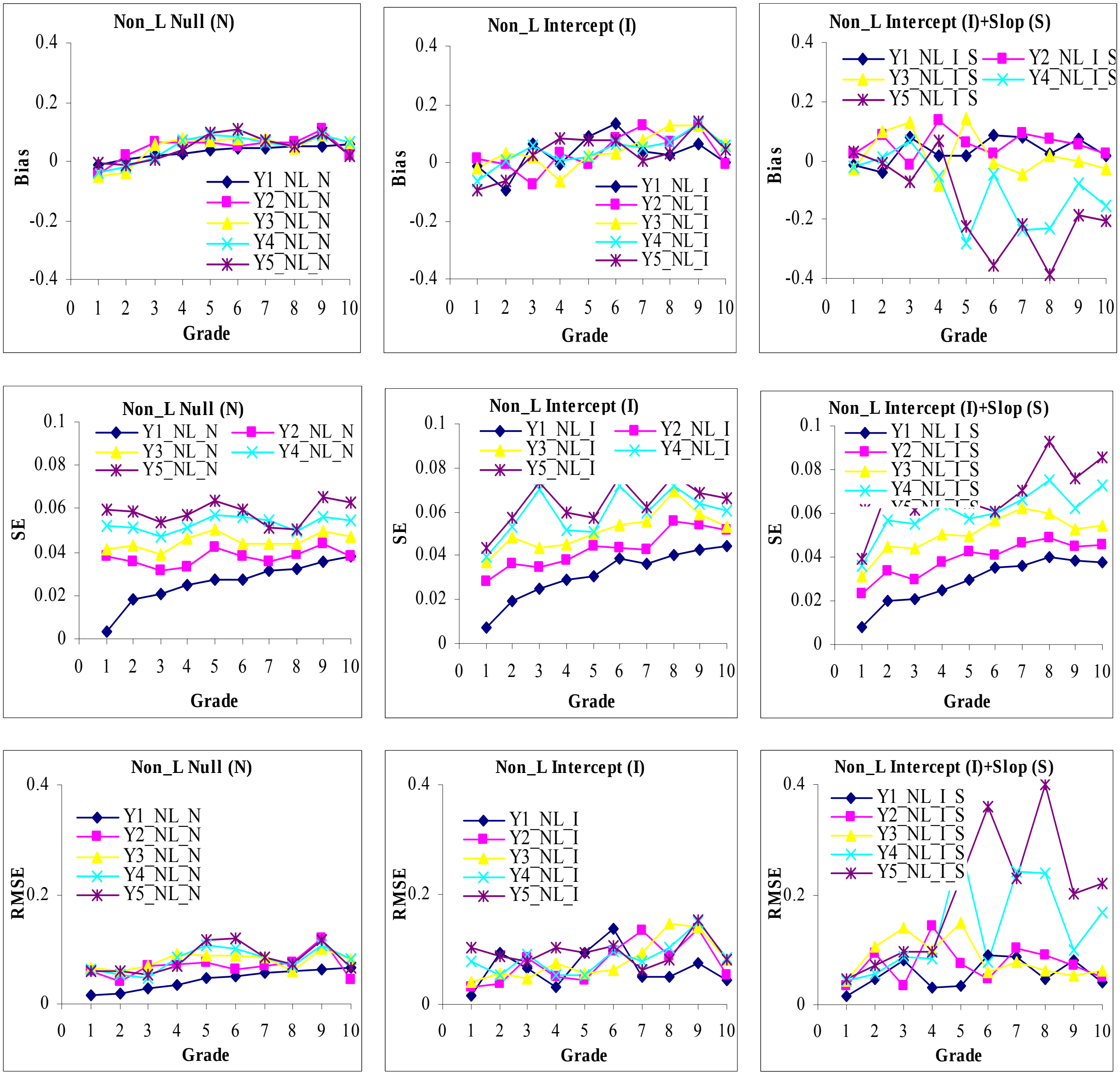


Figure 5. Bias, SE, and RMSE of Nonlinear Models (Null, Intercept, and Intercept + Slope) without Noise

Nonlinear Null Model

Nonlinear Intercept (With Noise)

Nonlinear Intercept + Slope (With Noise)

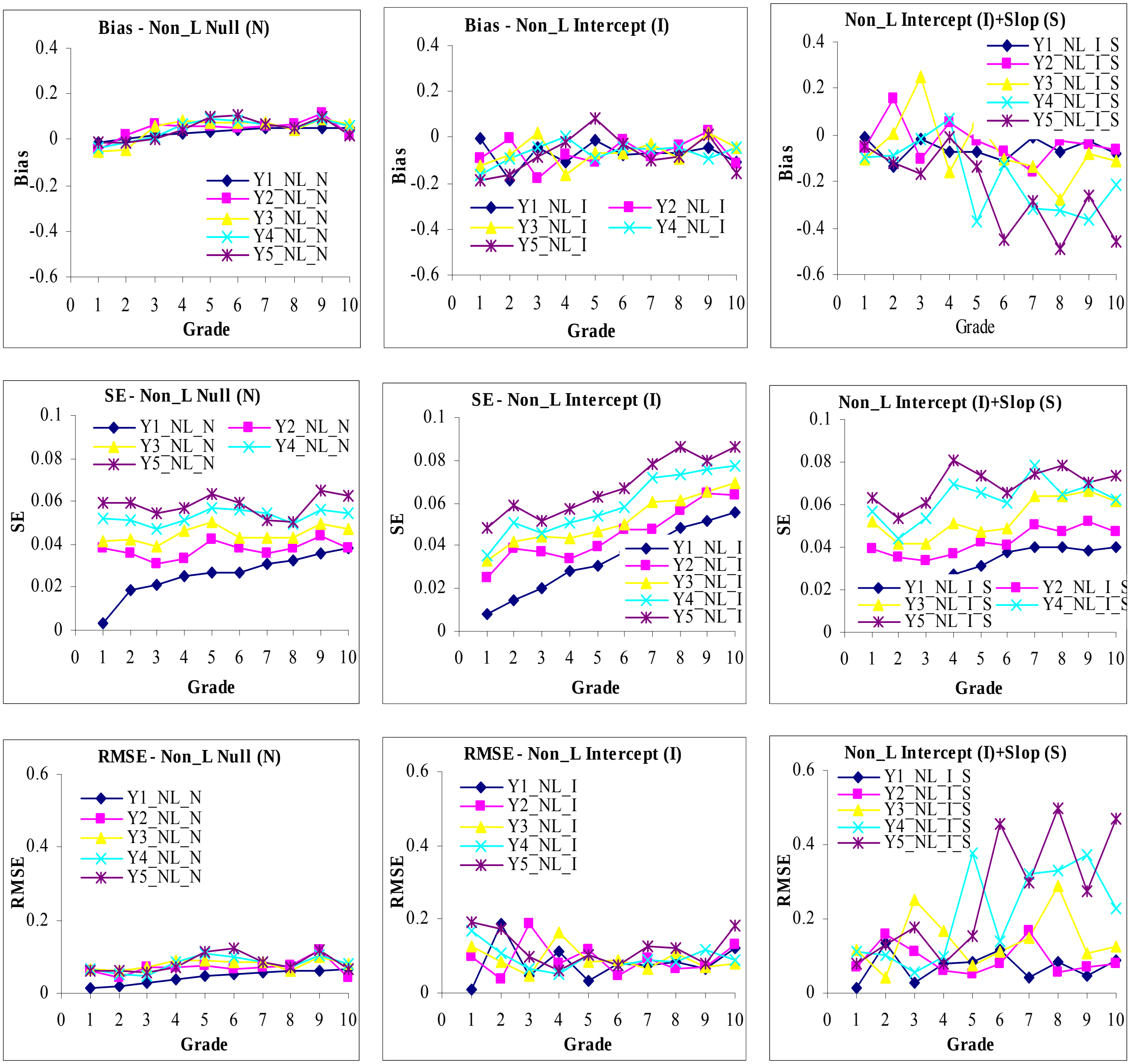


Figure 6. Bias, SE, and RMSE of Nonlinear Models (Null, Intercept, and Intercept + Slope) with Noise

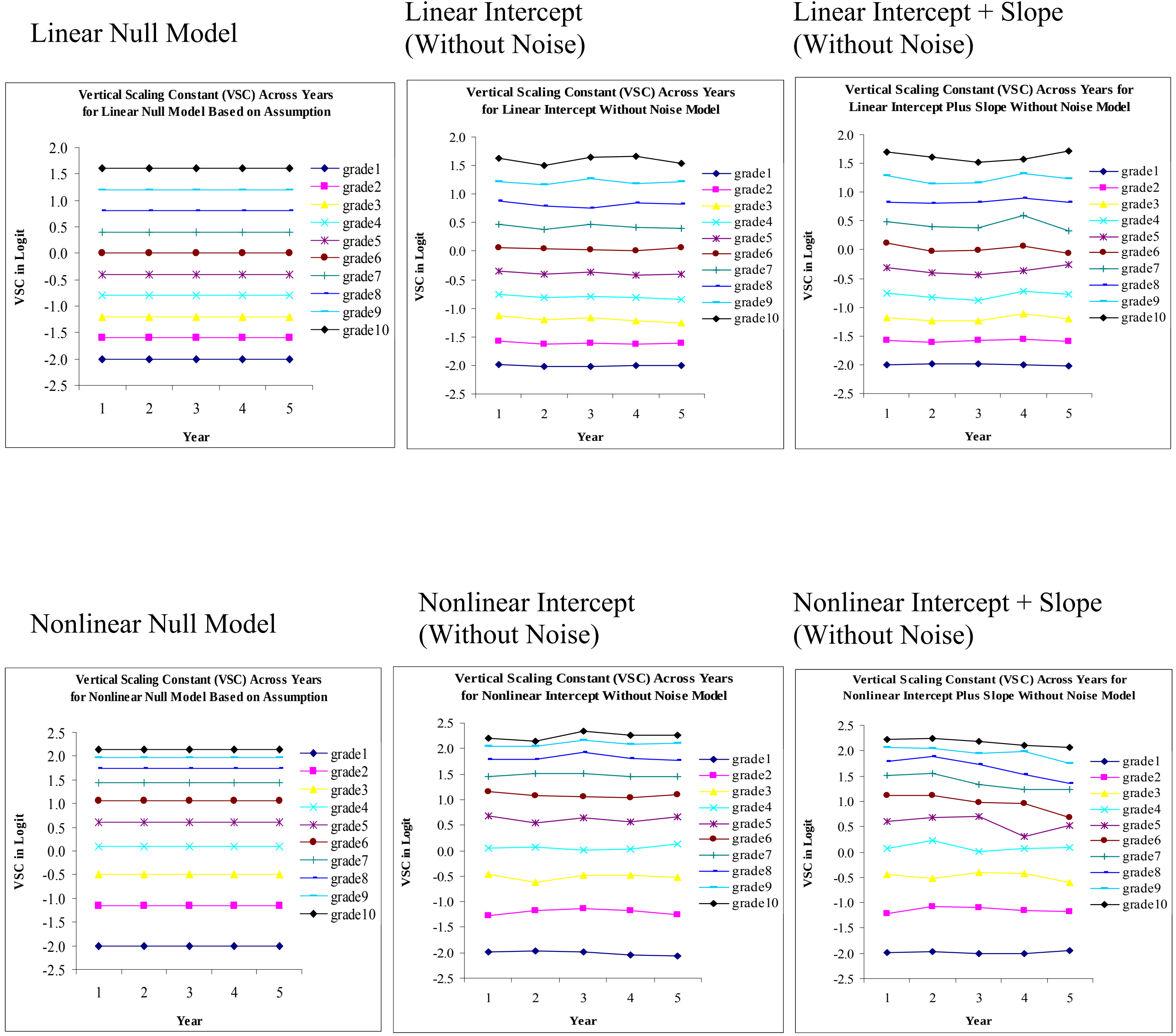
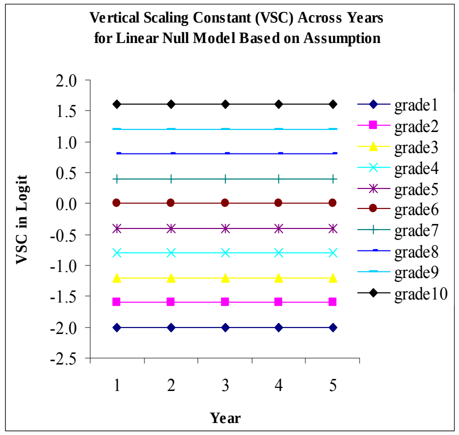
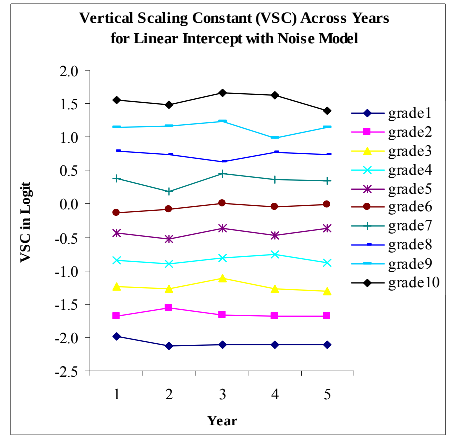


Figure 7. Vertical Scaling Constants of Both Linear and Nonlinear Models (Null, Intercept, and Intercept + Slope) without Noise

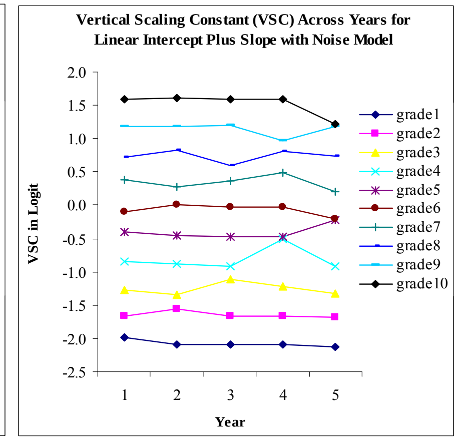
Linear Null Model



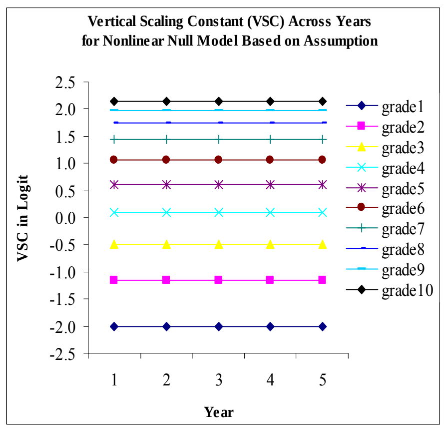
Linear Intercept (With Noise)



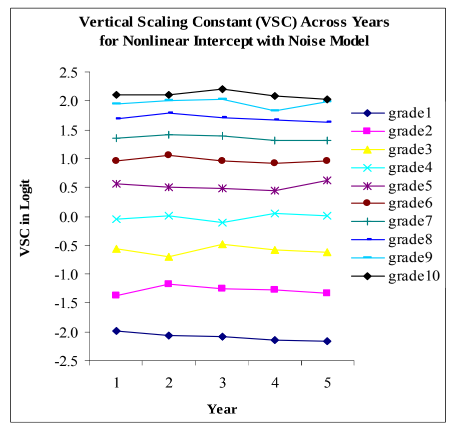
Linear Intercept + Slope (With Noise)



Nonlinear Null Model



Nonlinear Intercept (With Noise)



Nonlinear Intercept + Slope (With Noise)

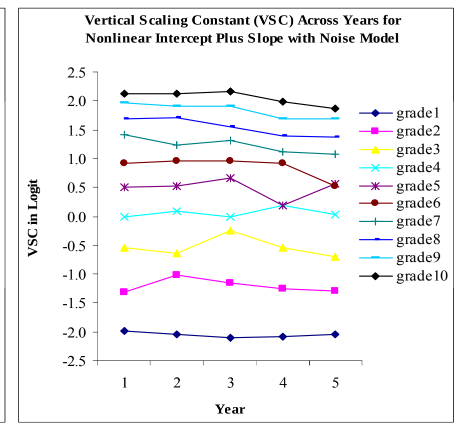


Figure 8. Vertical Scaling Constants of Both Linear and Nonlinear Models (Null, Intercept, and Intercept + Slope) with Noise

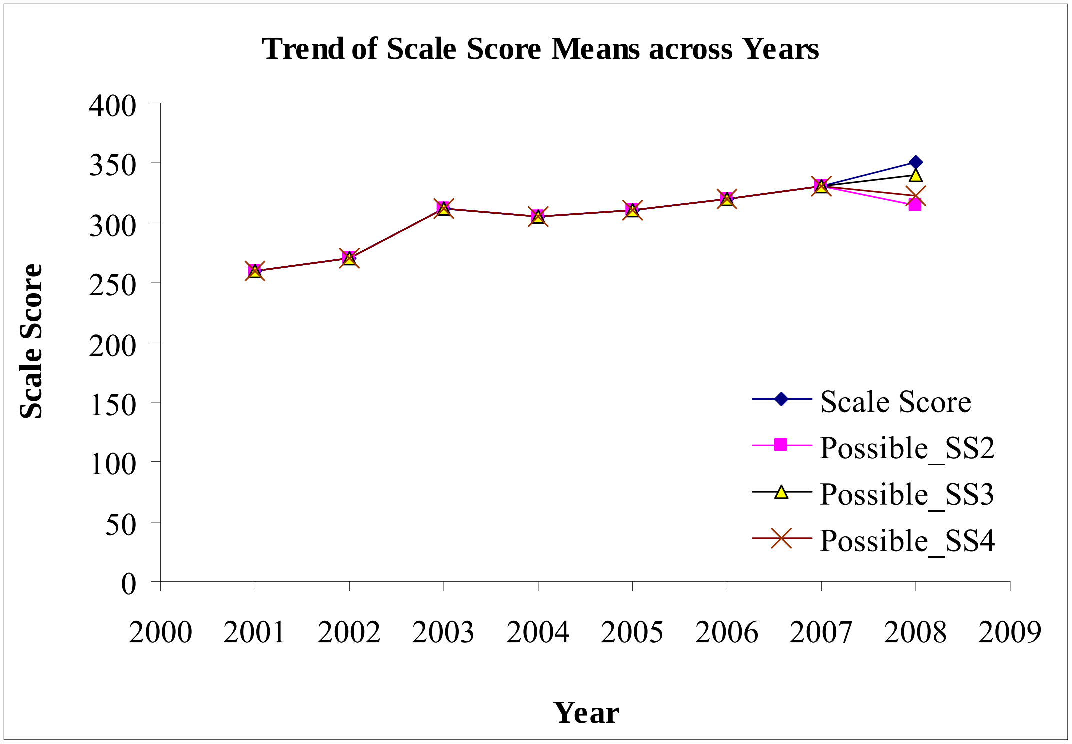


Figure 9. Example of Student Achievement Scores across Years