Psychological Bulletin
1987, Vol. 101, No. 1, 147-158

0033-2909/87/$00.75

Application of Hierarchical Linear Models to Assessing Change

Anthony S. Bryk
Department of Education
University of Chicago

Stephen W. Raudenbush
School of Education
Michigan State University

Recent advances in the statistical theory of hierarchical linear models should enable important breakthroughs in the measurement of psychological change and the study of correlates of change. A two-stage model of change is proposed here. At the first, or within-subject stage, an individual's status on some trait is modeled as a function of an individual growth trajectory plus random error. At the second, or between-subjects stage, the parameters of the individual growth trajectories vary as a function of differences between subjects in background characteristics, instructional experiences, and possibly experimental treatments. This two-stage conceptualization, illustrated with data on Head Start children, allows investigators to model individual change, predict future development, assess the quality of measurement instruments for distinguishing among growth trajectories, and to study systematic variation in growth trajectories as a function of background characteristics and experimental treatments.

Finding adequate measures of individual change and valid techniques for research on change are problems that have long perplexed behavioral scientists. Many concerns catalogued by Harris (1963) continue to trouble quantitative studies of psychological growth. On the substantive side, the most fundamental question is whether quantitative change over time is a meaningful issue. Lord (1963) and Bereiter (1963), for example, describe situations in which the structure of the abilities under study actually changes over the period of investigation. In such a case, the changing *structure* of abilities, not the *amount* of change, should be the prime research focus.

On the methodological side, it has been frequently noted (Linn & Slinde, 1977; Rogosa, Brand, & Zimowski, 1982) that high test stability (the correlation between scores across two or more time points) accompanies low change score reliability. Such low reliability may indicate that the measures are incapable of supporting precise statements about individual change. Yet when change score reliability improves, instrument stability typically declines, raising questions about whether the structure of the abilities themselves is changing.

Further, errors of measurement can produce particularly perverse effects on the assessment of change. Investigators routinely find, for example, that observed change over two occasions is negatively correlated with the subject's initial status. Bereiter (1963) demonstrated that this is, at least in part, a statistical artifact of measurement error. Even in situations in which the structural relation between change and initial status is positive, the observed relation can be negative (see, for example, Blomqvist, 1977). Thus, the true relation between initial status and rate of growth typically remains elusive.

The research reported here was supported by a grant from the Spencer Foundation. We also wish to acknowledge Mediax Interactive Technologies for their assistance in developing the computer programs.

Correspondence concerning this article should be addressed to Anthony S. Bryk, Department of Education, University of Chicago, 5835 Kimbark Avenue, Chicago, Illinois 60637.

These methodological problems have led to a bewildering array of well-intentioned but misdirected suggestions about the design and analysis of research on human change. Major reviews include Cronbach and Furby (1970), Linn and Slinde (1977), and Linn (1981). Recent research (Rogosa et al., 1982; Rogosa & Willett, 1983), however, dispells many of these misconceptions.

In essence, research on individual change has been plagued by inadequacies in conceptualization, measurement, and design. Brief reviews of these inadequacies follow.

1. Conceptualization. In any research context, a model of the phenomena under study is an important heuristic for guiding inquiry. Yet in most previous research on individual change, the model of individual growth is rarely addressed explicitly.
2. Measurement. Studies of change typically use tests that are developed to discriminate among individuals at a fixed point in time. Their adequacy for distinguishing the *rate of change* among individuals is rarely considered during the instrument design process. Further, statistical procedures routinely applied to these instruments, such as standardizing the scores to a common mean and variance over time, effectively eliminate the essence of individual growth (Rogosa et al., 1982). Psychometric procedures are needed that enable assessment of the adequacy of instruments for measuring both status and change.
3. Design. Much of the research on change has been based on data on individual status at two time points, for example, scores on a pretest and a posttest. In general, two time points provide an inadequate basis for studying change (Bryk & Weisberg, 1977; Rogosa et al. 1982). Further, even in instances in which data have been collected on multiple occasions, researchers have typically analyzed the data as a series of separate designs with two time points.

No coherent analytic strategy fully responds to these concerns. Recent developments in the statistical theory of hierarchical linear models (HLMs), however, now enable an integrated approach for studying the structure of individual growth, examining the reliability of instruments for measuring status and change, investigating correlates of status and change, and testing hypotheses about the effects of background variables and experimental interventions on individual growth. In subsequent sections we present a two-stage model of growth, discuss the statistical theory for estimating its parameters, and illustrate in detail its application using data on preschoolers' cognitive development.

Two-Stage Conceptualization

At Stage 1, each individual's observed development is conceived of as a function of an individual growth trajectory plus random error. This trajectory is determined by a set of individual parameters. At Stage 2, we assume that these individual parameters vary as a function of certain measurable characteristics of the individual's background and environment.

The explicit individual growth model at Stage 1 opens for scrutiny the theoretical basis of the study: psychologists can consider whether the abilities under investigation can plausibly be viewed as changing quantitatively over time. Further, the model's second stage requires a precise, falsifiable theory about how individual differences and experiences translate into differences in growth. This two-stage conceptualization implies the need for a model in which the parameters in the first stage become the outcome variables in the second stage. Because of this two-stage character, it is convenient to refer to this conceptualization of growth as a hierarchical linear model (HLM).

Within-Subject Model

In general, we assume that Y_{it}, the observed status of individual i at time t, is a function of a systematic growth trajectory or growth curve plus random error. It is convenient to assume that systematic growth over time can be represented as a polynomial of degree $K-1$. Thus, the within-subject model is

$$Y_{it} = \pi_{0i} + \pi_{1i}\, a_{it} + \pi_{2i}\, a_{it}^2 + \cdots + \pi_{K-1i}\, a_{it}^{K-1} + R_{it} \quad (1)$$

for $i = 1 \ldots n$ subjects, each of whom is observed on T_i occasions. Here a_{it} is the age of subject i at time t, π_{ki} ($k = 0, 1, \ldots K-1$) are the growth trajectory parameters for subject i, and R_{it} is the random error assumed normally distributed with a mean of zero and some covariance structure Σ_i. Note that Σ_i is dimensioned $T_i \times T_i$.

In general, Σ_i can take on a wide variety of structures, as will be discussed. We assume, however, that the within-subject error terms are uncorrelated across subjects, that is, $\text{cov}(R_{it}, R_{jt}) = 0$ for any values of t.

Between-Subjects Model

An important feature of Equation 1 is the assumption that the growth parameters (the π_{ki} values) vary across individuals. We formulate a between-subjects model to represent this variation. We are particularly interested in situations in which the individual growth parameters are a function of measured variables, such as characteristics of the individual's background (e.g., sex or social class) or of the experimental setting (e.g., type of curriculum, or amount of instruction).

Specifically, each of the k individual growth parameters can be modeled as

$$\pi_{ki} = \beta_{k0} + \beta_{k1}\, X_{k1i} + \beta_{k2}\, X_{k2i} + \cdots + \beta_{kP-1}\, X_{kP-1i} + U_{ki}, \quad (2)$$

where there are $p = 1, \cdots, P-1$ measured variables (X_{kp}), β_{kp} represents the effect of X_{kp} on the kth growth parameter, and U_{ki} is random error.

We further assume that the U_{ki} are normally distributed with mean zero and covariances given by

$$\text{cov}(U_{hi}, U_{ki}) = \text{cov}(\pi_{hi}, \pi_{ki}) = \tau_{hk} \quad (3)$$

for $h, k, = 0, 1, \cdots, K-1$.

In the language of analysis of variance, the parameters β_{kp} of the between-subjects model are known as fixed effects. The errors U_{ki} are the random effects—the unique increments to the growth parameters associated with each subject.

We note that each subject's growth can be measured at different ages and a different number of times. Thus, the within-subject model does not assume a uniform data collection design across subjects. The between-subjects model is also quite flexible. In particular, it can accommodate different X variables for each π_k.

Model Assumptions

To realize the increased flexibility of HLM requires careful attention to the necessary statistical assumptions. Three kinds of assumptions are needed: distributional assumptions, assumptions about the covariance structure, and assumptions about the metric in which the outcome variable is measured.

Distributional Assumptions

Both the individual outcomes, Y_{it}, and the growth parameters, π_{ki}, are assumed normally distributed. Because psychological measures often used as outcome variables have been developed intentionally to produce near-normal distributions, the first assumption is often not hard to satisfy. Further, analysts can check the validity of this assumption by examining histograms and normal probability plots.

The assumption that the growth parameters π_{ki} are normal is more difficult to assess because they are not directly observable. However, Waternaux, Laird, and Ware (1985) have recently developed methods for checking this normality assumption by comparing the sample frequency distribution of the growth parameter estimates (the π_{ki}^* estimates presented later) against the distribution expected under normality. The method identifies outliers and enables the investigator to assess their influence on substantive inferences.

Assumptions About the Covariance Structure Among the Observations

The HLM does not require the same data collection design for each individual. In the illustration that follows, some children have three data points, some four, and the spacing between the data collection points varies somewhat across cases. In such situations, it is often not sensible to assume a common covariance structure among the observations, as is customary in multivariate repeated measures (see, for example, Bock, 1975, p. 447ff). However, HLM is very flexible in that it permits modeling of a broad array of covariance structures through specifications imposed on both the individual growth model and on the random error term R_{it}.

Specification of the individual growth model. Specification of a model with random individual growth parameters as in Equation 1 has strong implications for the covariance structure among the observations. In general, both the variances of the Y_{it} and the covariances among them are functions of age (or time).

To illustrate, consider a simple within-subjects model:

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + R_{it}. \quad (4)$$

Here the intercept, π_{0i}, and the linear rate of growth, π_{1i}, determine the growth curve for each subject. For simplicity, suppose that information is available on just one background variable, X_i. The simplified between-subjects model is then given by

$$\pi_{0i} = \beta_{00} + \beta_{01}X_i + U_{0i}, \quad (5a)$$

and

$$\pi_{1i} = \beta_{10} + \beta_{11}X_i + U_{1i}. \quad (5b)$$

Combining Equations 4 and 5 yields a single linear model,

$$Y_{it} = \beta_{00} + \beta_{01}X_i + \beta_{10}\, a_{it} + \beta_{11}\, X_i\, a_{it} + e_{it}, \quad (6)$$

with

$$e_{it} = U_{0i} + U_{1i}\, a_{it} + R_{it}. \quad (7)$$

Equation 6 is a standard linear model with an intercept, β_{00}, and three predictors: the between-subjects variable, X_i; age, a_{it}; and the interaction term, X_ia_{it}. However, the error term, e_{it}, consists of two components: a component that depends on the random increments to the individual growth parameters, $U_{0i} + U_{1i}a_{it}$, and the random error, R_{it}.

Now suppose that the simplest assumption is used for R_{it} so that the errors R_{it} are independently distributed with a mean of zero and variance σ^2. Then, for any subject i, the errors e_{it} have variance

$$\mathrm{var}(e_{it}) = \tau_{00} + a_{it}^2\tau_{11} + 2a_{it}\tau_{01} + \sigma^2, \quad (8)$$

and covariance between any two observations at times t and t' of

$$\mathrm{cov}(e_{it}, e_{it'}) = \tau_{00} + (a_{it} + a_{it'})\,\tau_{01} + a_{it}a_{it'}\,\tau_{11}. \quad (9)$$

Equations 8 and 9 reveal the following statistical properties for this simple linear growth model. First, the variance of the observations is a function of age (or time), which is sensible, because individuals are presumed to grow at different rates. Second, each pair of observations for a given subject is correlated. Third, the size of this correlation depends on the spacing of the observations, on the relative magnitude of the variances among the intercepts and slopes, and on the covariance between them. A special case of Equations 8 and 9 merits particular mention. When the individual growth rates are constant across individuals (i.e., $\tau_{11} = 0$ and, by implication, $\tau_{01} = 0$), the covariance structure for the simple linear model reduces to a well-known structure, termed *compound symmetry,* often assumed in univariate repeated measures (Winer, 1971, p. 136).

In general, the structure of variances and covariances among the observations will depend both on the functional form assumed for the individual growth model and on the amount of variance and covariance among the individual growth parameters (see Bryk, 1977; Rogosa & Willett, 1985). By varying the specification of the individual growth model, it is possible to represent a broad range of covariance structures.

Specification of the random error term. It is most common with individual growth curve modeling (see Ware, 1985, p. 98) to assume a simple structure for the error term R_{it}, namely, the R_{it} are independently and normally distributed with a mean of zero and constant variance σ^2. Other assumptions can be accommodated, however. Although the statistical computations become more complex, Strenio (1981) presents an application in which the variance of R_{it} is person-specific, that is, $\mathrm{var}(R_{it}) = \sigma_i^2$. It is also possible to represent serial correlations among the random errors (Louis & Spiro, 1984), or to make R_{it} a function of measured characteristics such as age (Goldstein, 1986). These more complex models for R_{it} seem most useful when there are many time points per subject (Ware, 1985; Waternaux et al., 1985). For data sets with a short time series, the assumption of independent errors with constant variance is often most practical, and we use this assumption in the example that follows.

Metric of the Response Variable

Growth curve modeling requires that the outcome data collected at each time point be measured on a common metric, so that changes across time reflect growth and not changes in measurement scale. In our example we used item response theory to construct a common metric for each test, in logits, specifically to facilitate measurement of change.

Statistical Estimation

The statistical theory for estimating the parameters of HLM appears in a number of places under a variety of titles. The problem can be viewed as a mixed-model analysis of variance (Elston & Grizzle, 1962), as regression with random coefficients (Dielman, 1983; Rao, 1972; Rosenberg, 1973; Swamy, 1973), as James-Stein estimation (Efron & Morris, 1979), as a covariance components model (Harville, 1977), and as Bayesian estimation for linear models (Dempster, Rubin, & Tsutakawa, 1981; Lindley & Smith, 1972; Morris, 1983; Smith, 1973). We prefer the title "hierarchical linear model" because it highlights the class of substantive problems that can be addressed through these related approaches.

Estimation Assuming Known Variances

Within-subject model. To clarify the logic of statistical estimation for HLM without resorting to matrix notation requires a slight model simplification. Within subjects, we "center the data," so that the outcome, now denoted y_{it}, and age, a_{it}, are in deviation-score form. This eliminates the intercept from consideration and restricts our attention to the linear growth rate.[1] As a result, Equation 1 becomes

$$y_{it} = \pi_i a_{it} + R_{it}. \tag{10}$$

We also assume for simplicity that the errors R_{it} are independent, with common variance σ^2.

Under these assumptions, subject i's growth rate, π_i, can readily be estimated by means of ordinary least squares, based only on the repeated measurements for that subject.[2] The least squares estimate $\hat{\pi}_i$ is given by

$$\hat{\pi} = \sum a_{it}y_{it}/\sum a_{it}^2, \tag{11}$$

and the sampling variance of $\hat{\pi}_i$, for fixed π_i, is

$$v_i = \text{var}(\hat{\pi}_i \mid \pi_i) = \sigma^2/\sum a_{it}^2. \tag{12}$$

Between-subjects model. The simplified between-subjects model becomes

$$\pi_i = \beta_0 + \beta_1 X_i + U_i, \tag{13}$$

where the U_i are assumed to be independent and normally distributed with a mean of zero and variance τ. Now the outcome variable in Equation 13 is π_i, which is an unknown parameter. We have, however, an estimate for π_i from Equation 11. Our estimate $\hat{\pi}_i$ may be viewed as a function of the true π_i plus error, $\hat{\pi}_i - \pi_i$. Thus equation 13 can be rewritten as:

$$\hat{\pi} = \beta_0 + \beta_1 X_i + e_i, \tag{14}$$

where the $e_i = U_i + (\hat{\pi}_i - \pi_i)$ are independent and normally distributed with a mean of zero and variance $\tau + v_i = D_i$. Note that D_i is just the total variance in $\hat{\pi}_i$ and consists of the parameter variance, τ, and the sampling variance, v_i. That is,

$$D_i = \text{var}(\hat{\pi}_i) = \text{var}(\pi_i) + \text{var}(\hat{\pi}_i \mid \pi_i) = \tau + v_i. \tag{15}$$

Efficient estimates for β_0 and β_1 in Equation 14 can be obtained by using a weighted least squares regression where each subject's data ($\hat{\pi}_i$, X_i) are weighted inversely proportional to $\sqrt{D_i}$ (see Seber, 1977). We denote the resulting estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

Empirical Bayes estimation of individual growth curves. Estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ can be used to produce a second estimator of the individual growth parameter, π_i:

$$\hat{\hat{\pi}} = \hat{\beta}_0 + \hat{\beta}_1 X_i. \tag{16}$$

An interesting dilemma results from our having now derived two estimates of π_i ($i = 1, \ldots, n$), the individual growth rates for each subject. The first estimate, $\hat{\pi}_i$, from Equation 11, is simply the least squares estimate based on separate regressions for each subject. The second estimate, $\hat{\hat{\pi}}_i$, from Equation 16, is the predicted growth rate, based on X_i, the value of each subject's background variable. Rather than forcing a choice between one estimate and the other, empirical Bayes theory, reviewed by Morris (1983), provides a composite estimator, π_i^*, which is an optimally weighted average of $\hat{\pi}_i$ and $\hat{\hat{\pi}}_i$:

$$\pi_i^* = W_i\hat{\pi}_i + (1 - W_i)\hat{\hat{\pi}}_i. \tag{17}$$

Interestingly, the weights, W_i, have substantive interpretability. It can be shown (see Raudenbush & Bryk, 1985) that

$$W_i = \tau/(\tau + v_i), \tag{18}$$

Notice that W_i is just the ratio of the parameter variance in the growth rates, τ, to the total variance, $\tau + v_i$. This ratio is analogous to a reliability coefficient, where we compare the "true score" variance (in this case τ) to the observed score variance ($\tau + v_i$).

We see, then, that the influence of an individual's time series data (as captured by $\hat{\pi}_i$) on π_i^* depends on the reliability of the $\hat{\pi}_i$ estimate. When this is highly reliable, the HLM estimate for an individual's growth rate will lean heavily on the individual time series data. When this is unreliable, however, (i.e., $W_i \rightarrow 0$), π_i^* will be based primarily on the background data.

In essence, HLM uses whatever strength exists in the data in order to form its estimate for π_i. Both statistical theory and empirical research have shown that the composite estimator π_i^* has a smaller mean squared error than the estimator $\hat{\pi}_i$, based only on the individual data, or the estimator $\hat{\hat{\pi}}_i$, based only on the group data (Efron & Morris, 1979; Morris, 1983).

Finally, under the assumptions just laid out, π_i^*, $\hat{\beta}_0$, and $\hat{\beta}_1$ are all normally distributed with estimable sampling variances and covariances. These results provide the basis for both small and large sample hypothesis testing.

Estimation When the Variances Are Unknown

We have assumed so far that the variances σ^2 and τ were known. In most applications, however, those variance components must be estimated from the data. If all subjects are observed at identical times, variance estimation is straightforward. However, when the number and spacing of the time series observations vary across subjects, variance estimation requires iterative, numerical approaches that are just becoming accessible to researchers.

In particular, the development of the EM algorithm by Dempster, Laird, and Rubin (1977) affords a theoretically satisfactory and computationally manageable approach to variance/covariance estimation in HLMs. It has been successfully applied in a broad range of situations (see, for example, Dempster et al., 1981; Mason, Wong, & Entwhistle, 1983; Strenio, Weisberg, & Bryk, 1983), and we use it in the Illustrative Example.

Under fairly general conditions, the EM algorithm produces maximum likelihood estimates for variance components. These estimates have the desirable properties of being asymptotically unbiased, consistent, efficient, and asymptotically normally distributed. When the EM estimates are substituted for the unknown variances and covariances, the resulting β estimates are also maximum likelihood estimates with known asymptotic distributions. The latter provides the basis for large sample sta-

[1] The elimination of the intercept actually robs HLM of part of its strength. To the extent that the intercept and slope are correlated, HLM uses this association to increase estimation precision. Thus, centering the data is generally undesirable from a data analysis point of view, although it simplifies the presentation.

[2] If we assume a more general structure for Σ_i, then the first component to π_i^* would be derived from generalized least squares, where Σ_i^{-1} is used in the weight matrix.

Table 1
Enumeration of Head Start Sample

	Test							
Observations	Math		Reading		Perception		Natural science	
No. per subject								
3	12		84		45		14	
4	131		45		46		126	
Total *n*	143		129		91		140	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age at test								
1st	51.85	5.37	53.52	5.30	54.15	5.76	52.34	5.29
2nd	55.15	5.33	55.53	5.37	55.55	5.93	55.41	5.24
3rd	56.29	5.36	57.58	5.25	57.60	6.25	56.55	5.27
4th	58.35	5.36	58.85	5.30	58.67	5.98	58.65	5.26
Ability at test								
1st	−0.694	1.13	−0.480	0.827	−0.672	0.897	−0.731	0.808
2nd	−0.170	1.43	−0.310	0.761	−0.411	0.960	−0.218	0.940
3rd	−0.032	1.36	0.367	0.739	−0.269	0.794	−0.018	0.984
4th	0.303	1.14	0.200	0.843	0.219	1.14	0.407	1.82

tistical inference with HLM. The Appendix illustrates the logic of EM variance components estimation with HLM using the simple univariate model presented in Equations 10 and 13.

Generalization of the Model

A wide range of growth models can be formulated under HLM theory with EM variance component estimation. The within-subject model can be a polynomial of any degree. Subjects may be observed at different times and on a varied number of occasions. We can also assume a variety of models for the within-subject errors, R_{it}. Similarly, the between-subjects model may incorporate any number of background variables, and each between-subjects equation need not be identical for all growth parameters. For example, we might have one set of variables predicting π_{0i} and another predicting π_{1i}.

In the general case, when there is more than one individual growth parameter, the estimation formulas extend naturally. Whereas the univariate model discussed earlier required that we estimate the variance in the growth rate parameters, which we termed τ, and the sampling variance of the estimates $\hat{\pi}_i$, which we termed v_i, we now must estimate the variance/covariance matrices of the parameters and their errors of estimation, which we term $\mathbf{T}$ and $\mathbf{V}_i$. The elements of these two matrices are important substantively in estimating the reliability of measures and the correlation of change and initial status.

Illustrative Example[3]

We illustrate six potential uses of HLM in research on psychological change: describing the structure of the mean growth trajectory, estimating the extent of individual variation around mean growth, assessing the reliability of measures for studying both status and change, estimating the correlation between entry status and rate of change, examining how background and instructional variables influence change, and predicting future individual growth.

Data description. The sample consists of 143 preschoolers at three Head Start centers located in the Southeastern United States. The original design called for each child to be tested in four areas—reading, mathematics, language, and natural science—on each of four occasions approximately equally spaced throughout the year. In practice the testing dates varied across children and not every child was tested on all four occasions. Table 1 provides an enumeration of the sample, the ages (measured in months) at which data were collected, and descriptive statistics on the ability scores (measured in logits) for the four outcomes at each of the four occasions.

In addition to the test data, a limited amount of student background and program exposure information was collected. We use two such variables in our between-subjects model: "home language" (Spanish or English), and "amount of direct instruction" (hours per academic year). Nearly 15% of the children were from families in which Spanish was the dominant home language. The amount of direct instruction varied widely across centers and classrooms within centers. Both factors had significant bivariate relations with individual growth.

Examining the Individual Growth Model

Two distinct features of the growth system must be considered: the structure of the mean or average growth trajectory, and the nature of the deviations of the individual growth trajectories from the population mean. The first step in the analysis is to identify the degree of the polynomial to be fitted to the data. For clarity of exposition we consider only the simple linear individual growth model. The example is sufficiently rich, however, to demonstrate the statistical procedures used in estimating the model's parameters and examining its adequacy.

[3] All of the analyses described that follow were performed with an original FORTRAN program developed by the authors. The program is written in FORTRAN 77 and runs on a Hewlett-Packard 9000 mini-computer. The program is available for distribution by writing the authors.

Table 2
Estimated Mean Growth Parameters (Fixed Effects)

Dependent variable	*M* growth parameter estimates	*SE*	*Z*
Perception			
Intercept, β_{00}	−0.5095	0.0168	30.327*
Linear, β_{01}	0.1733	0.0256	6.770*
Natural science			
Intercept, β_{00}	−0.1352	0.0052	26.000*
Linear, β_{01}	0.1818	0.0249	7.274*

* $p < .001$.

Under a linear growth model, π_1 is the individual's growth rate over the data collection period and is identical to the expected amount of change that would occur in any fixed unit of time. The intercept parameter, π_0, is the true ability of each individual at some fixed time point. The specific time point depends on the scaling of the age metric. For the examples presented in this article, we defined the age metric in terms of the amount of time that had elapsed from the first data collection point. Under this specification, the π_0 parameter in the linear growth model represents the true ability level of the individual at the onset of data collection, or what we call the initial status. Thus, both π_0 and π_1 are of substantive interest.

Specifically, we pose the following within-subject model:

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + R_{it}, \qquad (19)$$

where we assume that the errors R_{it} are independent and normally distributed with common variance σ^2.

The between-subjects model is

$$\pi_{ki} = \beta_{k0} + U_{ki} \qquad (20)$$

for $k = 0, 1$.

Equation 19 assumes that for each individual, growth can be represented as a linear trajectory plus random error. Equation 20 is the simplest representation of a between-subjects model. We assume that each π_{ki} is random with some expectation, variance, and a covariance between them.

At this point our substantive interest focuses first on the β parameters, which represent the mean growth curve for the entire sample: β_{00} is the mean intercept, and β_{10} is the mean linear growth rate. We are interested in determining the simplest mean growth model that is consistent with the data. Thus we test the hypotheses

$$H_{0k}: \beta_{k0} = 0$$

for $k = 0, 1$.

A simple Z test is available by computing the ratio of each estimate to its standard error. This ratio has a large sample unit normal distribution under H_{0k}. Table 2 presents the results for two of the outcome measures, perception and natural science. We find statistically significant results on both tests for both the intercept and growth rate parameters, which suggests that both parameters are necessary for describing the mean growth trajectory. (In many applications, the intercept will be retained in the model even if it is not significantly different from zero.) The results for the reading and mathematics tests are not presented here for reasons discussed in the next section.

Next, we consider the nature of the deviations of the individual growth trajectories from the mean curve. The EM algorithm yields estimates for τ_{kk}, the variances of individual growth parameters, that is,

$$\tau_{kk} = \mathrm{var}(\pi_{ki}),$$

for $k = 0, 1$. These statistics are reported in Table 3.

We might wish to consider the hypotheses of no parameter variation among either initial status, π_0, or the growth rates, π_1. If the hypotheses that all children shared the same entry ability (i.e., $H_{00}: \tau_{00} = 0$) were true, the variation among the estimated $\hat{\pi}_{0i}$ would consist only of sampling variance, denoted v_{0i}. Thus, under the assumptions specified previously the statistic

$$\sum_i (1/v_{0i})(\hat{\pi}_{0i} - \hat{\beta}_{00})^2$$

would have a chi-square distribution with $n - 1$ degrees of freedom.

For our data, the test statistic equals 356.90 ($df = 139$, $p < .001$) and 432.64 ($df = 90$, $p < .001$) for the natural science and perception tests, respectively, which leads us to reject the null hypotheses and conclude that for both tests, children vary significantly in their initial status.

Similarly, we can examine the hypothesis that there are no individual differences among children's growth rates (i.e., $H_{01}: \tau_{11} = 0$). Retention of this hypothesis implies that the linear effect of age is fixed, with no subject-specific deviations. As noted earlier, retention of the hypothesis is equivalent to assuming compound symmetry of the variance–covariance matrix of the observations Y_{it}.

Under this hypothesis and the assumptions specified previously, the statistic

$$\sum_i (1/v_{1i})(\hat{\pi}_{1i} - \hat{\beta}_{10})^2$$

has a chi-square distribution with $n - 1$ degrees of freedom. Here the test statistic equals 724.91 ($df = 139$, $p < .001$) for natural science and 465.05 ($df = 90$, $p < .001$) for perception, which leads us in both cases to reject the hypothesis of no variation in linear growth rates (Table 3). Thus there is significant variation among individual growth rates, which also implies that the compound symmetry assumption is untenable for these data.

The procedures just illustrated generalize directly to more complex growth models. In principle, a polynomial of any degree can be fitted and tested as long as the time series is sufficiently long.[4] Transformation of the age metric, for example, by representing Y_{it} as a function of $\log(a_{it})$, can also be easily accomodated by transforming the age variable first, and then proceeding as shown previously. We suggest visual examination of the individual time series and mean trajectories to identify

[4] The statistical estimation procedures used in HLM do not require that the number of observations per individual exceed the number of parameters estimated per case (see Braun, Jones, Rubin, & Thayer, 1983). Although this offers interesting data analysis options, the practical utility of fitting multiparameter models from sparse data has not been adequately assessed.

Table 3
Estimated Variance Components (Random Effects)

Dependent variable	Estimated parameter variance, var(π)	χ^2	*df*
Perception			
Intercept, τ_{00}	0.6393	432.64*	90
Linear, τ_{11}	0.0690	465.05*	90
Natural science			
Intercept, τ_{00}	0.5111	356.90*	139
Linear, τ_{11}	0.0230	724.91*	139

* $p < .001$.

possible models that might be fitted to the data. In general, the mean growth curve and the individual growth curves could have different forms. For example, in fitting a quadratic model to the data, we might find that some individual trajectories with positive curvatures cancel out others with negative curvatures. In this case, a line would be a fine description for the group development, but an inadequate representation for individual growth.

Reliability of Assessments of Initial Status and Change

Before we expand the between-subjects model in order to examine the possible factors associated with the entry ability and growth rate, we should consider the psychometric characteristics, particularly the reliability, of the $\hat{\pi}$ estimates. If most of the variability in $\hat{\pi}$ were due to error, we would likely not find any systematic relations between these estimates and the second stage variables. We might then falsely conclude that there are no relations when in fact the data are incapable of detecting such relations.

Recall that for each of the K individual growth parameters the observed variance in the estimated individual parameters consists of sampling variance and parameter variance. Specifically,

$$\text{var}(\hat{\pi}_{ki}) = \text{var}(\hat{\pi}_{ki}|\pi_{ki}) + \text{var}(\pi_{ki}). \tag{21}$$

Following classical measurement theory, the ratio of the "true" parameter variance, var(π_{ki}), to the "total" observed variance, var($\hat{\pi}_{ki}$), is the reliability of the individual data estimate, $\hat{\pi}_{ki}$, as a measure of the "true" growth parameters, π_{ki}. Formally,

$$\rho_{ki} = \text{var}(\pi_{ki})/\text{var}(\hat{\pi}_{ki}) \tag{22}$$

for the $k = 0, \ldots, K-1$ growth parameter estimates, where ρ_{ki} is the reliability of individual i's estimate for growth parameter, π_{ki}.

Estimation of ρ_{ki} is straightforward. Because HLM provides maximum likelihood estimates for var(π_{ki}) and var($\hat{\pi}_{ki}$), substituting these estimates into Equation 22 provides maximum likelihood estimates for the individual growth parameter reliability coefficients. Averaging the estimates across the n individuals provides a summary index of the instrument's reliability in measuring each of the K growth parameters on this population of subjects. Table 4 shows the estimated variance components and instrument reliabilities for the four measures.

All four tests seem quite reliable as measures of entry status. The reliability of π_0 ranges from .65 for reading to .86 for perception. As measures of change, however, only perception and natural science demonstrate high reliability, .76 and .80, respectively (Table 4). The reliability of $\hat{\pi}_1$ for mathematics is only .03 and for reading, .24. These results indicate that there is very little variation in the growth rate parameters for the mathematics test and only slightly more on the reading test. To be sure, the abilities of individual students are developing in these areas (see Table 1), but the rate of development is relatively constant across individuals. As a result, the reliability of the growth rates is low for these two tests, and we have excluded them from further analysis.

Relation of Change to Initial Status

As noted at the beginning of the article, the correlation between change and initial status is an interesting characteristic of any collection of growth trajectories. It is impossible to estimate this relation on the basis of a simple pretest–posttest design. With multiwave data, however, HLM readily produces an estimate for this structural parameter as well. Under a linear individual growth model, the true correlation between change and initial status is just the correlation between π_0 and π_1. This correlation is a simple function of the covariances among the πs, that is,

$$\text{corr}(\pi_0, \pi_1) = \tau_{01}/(\tau_{00} \cdot \tau_{11})^{1/2}. \tag{23}$$

Because the EM algorithm provides maximum likelihood estimates for each of the elements in Equation 23, substituting these estimates into Equation 23 yields a maximum likelihood estimate for corr(π_0, π_1). For the perception data, the estimated correlation of change with initial status is −.562 and for natural science, it is −.278.

Thus, HLM methods enable us to infer that the negative correlation between estimated entry status and growth rate is not entirely spurious. A substantial part of the observed negative correlation is attributable to the negative association of the parameters themselves. In general, the students whose initial status going into Head Start is low tend to grow at a somewhat faster rate while in the program.

Correlates of Change and Status

Another major application of HLM focuses on the relations between measured background and program characteristics on

Table 4
Reliability of Initial Status, and Growth Rates Estimates

Dependent variable	Estimated parameter variance	Estimated total variance	Reliability[a]
Initial status, π_{0i}	var(π_{0i})	var($\hat{\pi}_{0i}$)	var(π_{0i})/var($\hat{\pi}_{0i}$)
Math	0.584	0.738	.791
Reading	0.309	0.477	.648
Perception	2.237	2.59	.862
Natural science	1.689	1.98	.854
Growth rates, π_{1i}	var(π_{1i})	var($\hat{\pi}_{1i}$)	var(π_{1i})/var($\hat{\pi}_{1i}$)
Math	.0007	.0207	.034
Reading	.0027	.0113	.240
Perception	.0314	.0415	.756
Natural science	.0400	.0510	.799

[a] Estimated proportion of total variance that is parameter variance.

the one hand, and entry status (π_0) and growth rate (π_1) on the other. In illustrating this with the natural science data, we assume a linear growth model and introduce two variables in the between-subjects model: X_1 is the home language (1 = Spanish, 0 = English), and X_2 are the hours of direct classroom instruction (a continuous variable). Thus, our within-subject model is

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + R_{it}, \qquad (24)$$

and our between-subjects model is

$$\pi_{0i} = \beta_{00} + \beta_{10}X_{1i} + \beta_{20}X_{2i} + U_{0i}, \qquad (25)$$

and

$$\pi_{1i} = \beta_{01} + \beta_{11}X_{1i} + \beta_{21}X_{2i} + U_{1i}. \qquad (26)$$

Table 5 presents selected results from this analysis. Because data in this example are nonexperimental, causal inferences should be approached with caution. The discussion that follows is intended primarily to illuminate the technical interpretation of the results, rather than to make specific substantive inferences about preschool education.

We first consider the estimates for the fixed effects. Neither home language nor hours of instruction is significantly related to entry ability. Nevertheless, the direction of the estimated effects seems plausible. On average, Spanish speakers start behind English speaking children by .463 logits, that is, $\hat{\beta}_{10} = -.463$. It is a commonly encountered phenomenon in Head Start that children from non-English speaking families tend to score lower initially, but are also likely to show rapid progress. The positive relation between total hours of instruction and initial status is also reasonable because the first testing occasion, t_1, occurred between 6 and 14 weeks into the program year. Because a substantial amount of instruction had already been given, the observed effect is not surprising.

Table 5
Effects of Home Language and Hours of Instruction on Growth Parameters for Natural Science

Effects	Coefficient	*SE*	*Z*
Fixed, β			
Effect of home language on initial status, $\hat{\beta}_{10}$	−0.463	0.304	−1.52
Effect of hours of instruction on initial status, $\hat{\beta}_{20}$	1.523×10^{-3}	0.853×10^{-3}	1.79
Effect of home language on rate of change, $\hat{\beta}_{11}$	0.187	0.045	4.20*
Effect of hours of instruction on rate of change, $\hat{\beta}_{21}$	4.735×10^{-4}	1.252×10^{-4}	3.78*
Random, π			
Total variance in growth parameters estimates			
Initial status, var($\hat{\pi}_0$)	1.978		
Growth rate, var($\hat{\pi}_1$)	0.051		
Unconditional parameter variance			
Initial status, var(π_0)	1.689		
Growth rate, var(π_1)	0.041		
Reduction in parameter variance due to X_1, X_2			
Initial status, var(π_0) − var($\pi_0 \mid X_1, X_2$)	.928		
Growth rate, var(π_1) − var($\pi_1 \mid X_1, X_2$)	.031		
R^2: Percentage of total variance explained			
Initial status	38.4		
Growth rate	56.0		
R^2: Percentage of parameter variance explained			
Initial status	54.9		
Growth rate	75.0		

* $p < .001$.

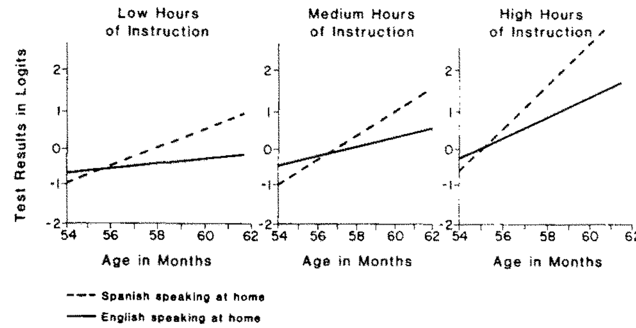

Figure 1. Estimated mean growth trajectories incorporating information on home language and hours of instruction (natural science test). (The results presented in this figure are based on a between-subjects model including both home language [Spanish vs. English] and hours of instruction [M = 319.7, SD = 123.03]. The relation between age and test results was plotted with hours of instruction held constant at three values: 196.7 [low], 319.7 [medium], and 442.7 [high]. The low and high values are 1 SD below and above the mean, respectively.)

Both home language and hours of instruction relate significantly to individual growth rates. The scores for children whose home language is Spanish are increasing, on average, at a rate .187 logits per month faster than the scores of their English-speaking companions (holding constant hours of instruction). Similarly, each additional hour of instruction per year is associated with a .0004735-logit increment to the growth rate (holding constant home language). To understand the latter result, consider the expected growth rates for two children who have the same home language but varying amounts of instruction. Specifically, suppose the first child receives 40 hr per month of instruction and the second 80 hr. (These numbers approximate the minimum and maximum hours of instruction in the Head Start sample.) The model predicts that over a 9-month period, the extra 40 hr per month of instruction received by the second child will yield an increment to that child's growth rate of 9 × 40 × .0004735, or .170 logits per month. That is, the child in the 80 hr per month program will be expected to grow at a rate of .170 logits per month faster than his counterpart in the 40 hr per month program.

These relations between background variables and π_0 and π_1 are illustrated in Figure 1. Estimated mean growth trajectories for children of English and Spanish backgrounds are plotted separately for selected values of hours of instruction. In each of the figure's panels, the Spanish speakers start out behind, but grow at a faster rate than the English speakers. As the amount of instruction increases, the expected growth rates for both groups

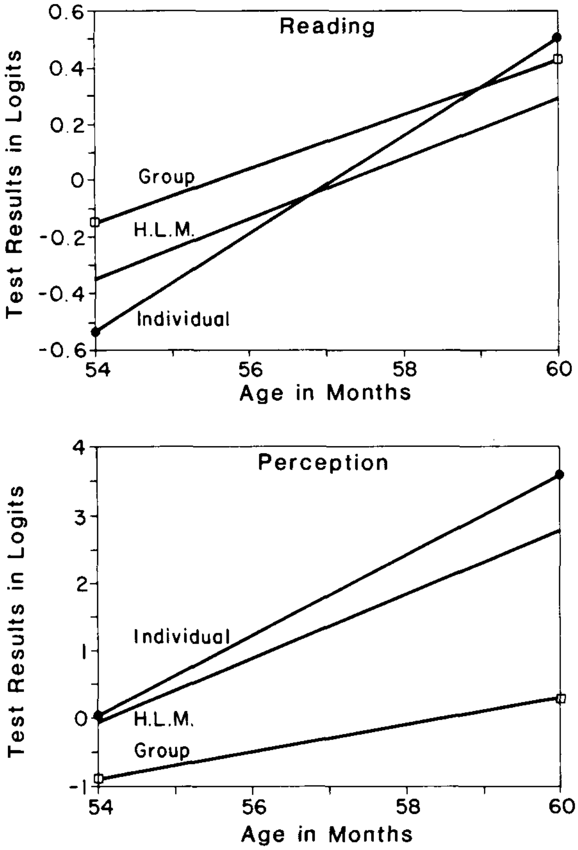

Figure 2. Alternative individual growth trajectory estimates from the hierarchical linear model (HLM).

increase, with the greatest progress being displayed in the third panel, where the hours of instruction are highest.

Returning to Table 5, estimates are also provided for the variances associated with the random effects, π_{0i} and π_{1i}, in the model. In considering the explanatory power of a HLM model, two different versions of R^2 can be computed: the percentage of total variance explained, or the percentage of the parameter variance explained. The latter is more informative because a part of the total variance is sampling error in $\hat{\pi}_k$, which by definition is not explainable by background variables. If this sampling variance is substantial, the percentage of total variance explained may be very small even when the model is accounting for most of the "explainable variance," that is, the parameter variance. In our illustration, for example, home language and hours of instruction account for 54.9% of the parameter variance in the initial status, and 75% of the parameter variance in growth rates on the natural science test.

Using HLM Estimates to Improve Predictions About Individual Growth

Recall from Equation 16 that HLM produces for each subject an estimate of each individual growth parameter, π_{ki}, which is a weighted combination of two estimates: one derived solely from subject i's own time series data, and one derived from knowing the relation between background variables and growth parameters. As noted previously, the composite estimator, π^*, has smaller mean square error and as a result provides a better basis for predictions (both in terms of interpolation and extrapolation) than either of its component parts.

Figure 2 shows the estimated growth trajectories on the reading and perception tests for the first subject in our data set. Both the HLM composite estimator and its component trajectories are presented. These results illustrate the important role that the reliability of the individual growth parameter estimates, $\hat{\pi}$, play in determining π^*. Recall from Table 4 that whereas the reliability of the intercept parameter estimates for reading was modest, .648, the growth rate reliability was quite low, .240. As a result, the HLM estimate for the intercept for Subject 1 is partly between the group and individual estimates. The HLM growth rate estimate, however, is based almost entirely on the background data. This is sensible, given the low reliability of the individual growth data estimates. The estimates for the perception test, however are quite different. Because the individual data estimates for both intercept and growth rates are reliable on this measure, the HLM estimates more closely follow these results.

In short, HLM capitalizes on any strength in the data. If the individual growth trajectory estimates are reliable then HLM weights them heavily. If the latter estimates are not reliable, the model substitutes values from mean growth trajectories that are conditioned on available background information. Finally, HLM also provides theory for approximate confidence interval estimates for growth predictions (see Strenio et al., 1983).

Summary and Discussion

Long-standing problems of measurement, conceptualization, and design have beset research on psychological change. However, developments over the past 10 years in the statistical theory of HLMs now enable an integrated approach for (a) studying the structure of individual growth and estimating important statistical and psychometric properties of collections of growth trajectories; (b) discovering correlates of change, that is, factors that influence the rate at which individuals develop; and (c) testing hypotheses about the effects of one or more experimental or quasi-experimental treatments on growth curves.

The approach is based on a two-stage, hierarchical model. At the first stage, a within-subjects model expresses status on a given trait as a function of an individual growth trajectory plus random error. At the second or between-subjects stage, individual growth parameters vary as a function of differences between subjects in background and experience.

An example based on Head Start data illustrated key analytic uses of HLM: (a) describing the structure of the mean growth trajectory; (b) estimating the extent and character of individual variation around mean growth; (c) assessing the reliability of measures for studying both status and change; (d) estimating the correlation between subjects' entry status and rates of growth; (e) estimating correlates of both status and change; (f) assessing the adequacy of between-subjects models by estimating reduction in unexplained parameter variance (reduction in uncertainty about the individual growth parameters as distinguished from errors in their estimation); and (g) predicting future individual growth.

Although we do not illustrate it, HLM can be applied in experimental and quasi-experimental settings. An experimental design could be incorporated into the between-subjects model to represent both treatment group membership and possibly covariates. (Experimental manipulations such as a treatment-interrupted time series could also be incorporated into the within-subject model.) This flexibility encourages an expanded conceptualization for the effects of treatments on individuals. Whereas experimental research conventionally assumes that a treatment adds a constant increment to each individual's value on the outcome variable, HLM permits a broader representation of the effects of interventions on the structure of growth, including (for example) effects on growth rates, on the correlation of entry status and growth, on the shape of the curvature of learning curves, and on the variability of growth trajectories.

The HLM approach requires multi-time point data. This simply reflects the reality that adequate design for the study of individual change generally requires more than two time points. In its handling of these data, however, HLM is quite flexible in that the number and timing of observations may differ across subjects.

The special strengths of HLM in individual prediction are noteworthy. In making predictions, the model draws on whatever strengths are available in the data: if within-subject data are precise, the model weights that data heavily. If between-subjects relations are strong, that data receives emphasis. If the growth parameters are correlated, empirical Bayes exploits this too, although presentation of the model in matrix notation is required to demonstrate this benefit (see Strenio et al., 1983).

The study of growth curves using HLM requires special care to distributional assumptions, covariance assumptions, and the metric of measurement. More research is needed on the consequences of violating the assumption of normality of the growth parameters. However, such violations will clearly have relatively little influence on estimates of the fixed effects (the βs). These estimates are based on generalized least squares theory, a defensible choice without resort to distributional assumptions. Also, the π^* estimates of individual growth parameters will typically be sensible even when the normality assumption fails.

More problematic are inferences based directly on the estimated variances and covariances, as these estimates depend more heavily on the normality assumption and are also likely to be imprecise when sample sizes are small. This means that the estimated correlation between initial status and rate of change and the estimated reliability of the growth parameters should be regarded as tentative when normality is questionable or if samples are small. More research is also needed on the robustness of these estimates to nonnormality and on the sample sizes needed for stable estimation. In addition, hypothesis testing leans more heavily on distributional assumptions than does point estimation of parameters. As a result, statistical inferences should be regarded as approximate when distributional assumptions are in question.

On balance, hierarchical linear models seem broadly applicable to the study of change and are likely to extend substantially the empirical research on change. To the extent that HLM enriches the class of testable hypotheses about the structure of growth, it may also encourage a broadened discussion about the nature of change itself.

References

Bereiter, C. (1963). Some persisting dilemmas in the measurements of change. In C. W. Harris (Ed.), *Problems in the measurement of change* (pp. 3–20). Madison: University of Wisconsin Press.

Blomqvist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association, 72,* 746–749.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill.

Braun, H. I., Jones, D. H., Rubin, D. B., & Thayer, D. T. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. *Psychometrika, 48,* 171–181.

Bryk, A. S. (1977). An investigation of the effects of alternative statistical adjustment strategies in the analysis of quasi-experimental growth data (Doctoral dissertation, Harvard University Graduate School of Education, 1976). *Dissertation Abstracts International, 38,* 3761B.

Bryk, A. S., & Weisberg, H. I. (1977). Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin, 84,* 950–962.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"—or should we? *Psychological Bulletin, 74,* 68–80.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–8.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association, 76,* 341–353.

Dielman, T. E. (1983). Pooled cross-sectional and time-series data: A survey of current statistical methodology. *Journal of the American Statistical Association, 78,* 111–122.

Efron, B., & Morris, C. (1979). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association, 74,* 311–319.

Elston, R. C., & Grizzle, J. E. (1962). Estimation of time response curves and their confidence bands. *Biometrics, 18,* 148–159.

Goldstein, H. I. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73,* 43–56.

Harris, C. W. (Ed.). (1963). *Problems in the measurement of change.* Madison: University of Wisconsin Press.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association, 72,* 320–338.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B, 34,* 1–41.

Linn, R. L. (1981). Measuring pretest–posttest performance changes. In R. Berk (Ed.), *Educational evaluation methodology: The state of the art* (pp. 84–109). Baltimore, MD: Johns Hopkins University Press.

Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research, 47,* 121–150.

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 21–38). Madison: University of Wisconsin Press.

Louis, T. A., & Spiro, A., III. (1984). *Fitting first order auto-regressive models with covariates* (Tech. Rep.). Cambridge, MA: Harvard University School of Public Health, Department of Biostatistics.

Mason, W. M., Wong, G. M., & Entwistle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 72–103). San Francisco: Jossey-Bass.

Morris, H. S. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association, 78,* 47–65.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association, 67,* 112–115.

Raudenbush, S. W. (1984). Applications of a hierarchical linear model in educational research (Doctoral dissertation, Harvard University Graduate School of Education, 1983). *Dissertation Abstracts International, 45,* 1713A.

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10,* 75–98.

Rogosa, D. R., Brand, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90,* 726–748.

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of difference scores in the measurement of change. *Journal of Educational Measurement, 20,* 333–343.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Rosenberg, B. (1973). Linear regression with randomly dispersed parameters. *Biometrika, 60,* 65–72.

Seber, G. A. F. (1977). *Linear regression analysis.* New York: Wiley.

Smith, A. F. M. (1973). A general Bayesian linear model. *Journal of the Royal Statistical Society, Series B, 35,* 61–75.

Strenio, J. F. (1981). *Empirical Bayes estimation of linear model parameters.* Doctoral dissertation, Harvard University Department of Statistics.

Strenio, J. L. F., Weisberg, H. I., & Bryk, A. S. (1983). Empirical Bayes estimation of individual growth curves parameters and their relationship to covariates. *Biometrics, 39,* 71–86.

Swamy, P. A. V. (1973). Criteria, constraints, and multi-collinearity in random coefficient regression models. *Annals of Economic and Social Measurement, 2,* 429–450.

Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *Journal of the American Statistical Association, 80,* 95–101.

Waternaux, C., Laird, N. M., & Ware, J. H. (1985). *Methods for analysis of longitudinal data: Blood lead concentrations and cognitive development* (Tech. Rep.). Cambridge, MA: Harvard University School of Public Health, Department of Biostatistics.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

Appendix

To illustrate the logic of variance components estimation, we consider a simple two-stage model. Within subjects,

$$Y_{it} = \pi_i a_{it} + R_{it},\ R_{it} \sim \mathrm{N}(0, \sigma^2). \tag{A1}$$

We assume for simplicity that y_{it} and a_{it} have means of zero, that is, that they have been centered around their population means.

Between subjects, we model the growth rates π_i as a function of their grand mean, μ_π, and the effect of a single predictor, X_i:

$$\pi_i = \mu_\pi + \beta(X_i - \bar{X}) + U_i,\ U_i \sim \mathrm{N}(0, \tau). \tag{A2}$$

This model is equivalent to the model

$$\hat{\pi}_i = \mu_\pi + \beta(X_i - \bar{X}) + e_i, \tag{A3}$$

where the $e_i = u_i + (\hat{\pi}_i - \pi_i)$ are independent and normally distributed with a mean of zero and variance $\tau + v_i = D_i$.

Estimation of σ^2, τ. If all subjects are observed at identical times, variance estimation is straightforward. Within subjects,

$$\hat{\sigma}^2 = \sum_i \sum_t (y_{it} - \hat{\pi}_i a_{it})^2/[n(T-1)], \tag{A4}$$

where there are T observations on each of n subjects. The estimated sampling variance of the estimated regression coefficients $\hat{\pi}_i$ would be identical for each case:

$$\hat{v}_i = \hat{v} = \hat{\sigma}^2/\sum a_{it}^2. \tag{A5}$$

The estimated variance of $\hat{\pi}_i$ across all subjects would be

$$\hat{D}_i = \hat{D} = \sum_i [\hat{\pi} - \hat{\mu}_\pi - \hat{\beta}(X_i - \bar{X})]^2/(n-2). \tag{A6}$$

As a result, an unbiased, efficient estimate of τ follows directly:

$$\hat{\tau} = \hat{D} - \hat{v}. \tag{A7}$$

When the number and/or spacing of the time series observations vary across subjects, an iterative method, like the EM algorithm, is required to estimate variances.

Logic of EM. As we indicate in the article, β, μ_π, and π_i ($i = 1, 2, \ldots n$) may be estimated when both σ^2 and are τ known. These estimates are maximum likelihood estimates (Raudenbush, 1984). Alternatively, if β, μ_π, and the π_i were known, maximum likelihood estimates of σ^2 and τ would follow:

$$\hat{\sigma}^2 = \sum_i \sum_t (y_{it} - \pi_i a_{it})^2/\sum T_i \tag{A8}$$

and

$$\hat{\tau} = \sum_i [\pi_i - \mu_\pi - \beta(X_i - \bar{X})]^2/n. \tag{A9}$$

Thus it is easy to derive maximum likelihood estimates for either the structural parameters (π_i, μ_π, and β) assuming the variances (σ^2 and τ) are known, or for the variances assuming the structural parameters are known. The troublesome part is deriving estimates for all of these parameters simultaneously. The fact, however, that the structural parameter estimates depend on the variances and the variance estimates in turn depend on the structural parameters provides the basis for the application of the EM algorithm to this problem.

Equations A8 and A9 are recognizable as residual mean squares. We are interested in determining the posterior expectations of these mean squares, when π_i, μ_π, and β are unknown, in terms of empirical Bayes estimates π_i^*, $\hat{\mu}_\pi$, and $\hat{\beta}$. It can be shown that

$$\mathrm{E}(\hat{\sigma}^2|\mathbf{y}) = \sum\sum (y_{it} - \pi_i^* a_{it})^2/\sum T_i + \sum\sum a_{it}^2 v_{\pi_i}^*/\sum T_i, \tag{A10}$$

where

$$v_{\pi_i}^* = W_i v_i + (1 - W_i)^2[v_{\hat{\mu}} + (X_i - \bar{X})^2 v_{\hat{\beta}}].$$

Similarly,

$$E(\hat{\tau}|\mathbf{y}) = \sum [\pi_i^* - \hat{\mu}_\pi - \hat{\beta}(X_i - \bar{X})]^2/n + \sum v_{\hat{U}_i}^*, \tag{A11}$$

where

$$v_{\hat{U}_i}^* = \mathrm{var}(U_i^*) = \mathrm{var}[\pi_i^* - \hat{\mu}_\pi - \hat{\beta}(X_i - \bar{X})]$$
$$= W_i v_i + W_i^2[v_{\hat{\mu}} + (X_i - \bar{X})^2 v_{\hat{\beta}}],$$

and $v_{\hat{\mu}}$ and $v_{\hat{\beta}}$ are the sampling variances under weighted least squares.

The first components in Equations A10 and A11 are identical to Equations A8 and A9 except that we have now substituted the empirical Bayes estimates for the parameters. Intuitively, we would expect that when π_i, μ_π, and β are unknown, Equations A8 and A9 (with estimates substituted) would be too small as expressions for $\hat{\sigma}^2$ and $\hat{\tau}$, as they fail to take into account the uncertainty associated with the estimation of π_i, μ_π, and β from the data. The second components in Equations A10 and A11 represent adjustments to the mean squares to account for this additional variability. The actual computations in applying the EM algorithm proceed as follows.

1. Initial values for σ^2 and τ must first be determined. A natural starting point is to compute ordinary least squares (OLS) regressions for each subject, yielding estimates, $\hat{\pi}_i$. The estimates, $\hat{\pi}_i$, are then regressed on X_i, also by means of OLS. The residual mean squares from both regressions yield first cut estimates of σ^2 and τ. For σ^2, the sums of squares from each within-subject model are pooled, as in Equation A4. For τ, a first cut estimate is similar to $\hat{\tau}$ of Equation A7, where $\hat{v}$ is the average value of $\hat{v}_i$.

2. These initial estimates of the variances are used to find initial estimates for π_i, μ_π, and β.

3. The structural parameter estimates from Step 2 are substituted into Equations A10 and A11 to derive new estimates for σ^2 and τ.

4. These new variance estimates are in turn used to generate new estimates for π_i, μ_π, and β.

5. The process in Steps 3 and 4 iterates back and forth between estimating variances and estimating structural parameters until convergence.

Dempster, Laird, and Rubin (1977) showed that the estimates of σ^2 and τ converge to maximum likelihood estimates. Generalization to more complex within- and between-subjects models proceeds naturally, but requires a general matrix formulation.

Received September 9, 1985
Revision received April 5, 1986 ■