Psychological Methods
1997, Vol. 2, No. 4, 371-402

1082-989X/97/$3.00

General Longitudinal Modeling of Individual Differences in Experimental Designs: A Latent Variable Framework for Analysis and Power Estimation

Bengt O. Muthén
University of California, Los Angeles

Patrick J. Curran
Duke University

The generality of latent variable modeling of individual differences in development over time is demonstrated with a particular emphasis on randomized intervention studies. First, a brief overview is given of biostatistical and psychometric approaches to repeated measures analysis. Second, the generality of the psychometric approach is indicated by some nonstandard models. Third, a multiple-population analysis approach is proposed for the estimation of treatment effects. The approach clearly describes the treatment effect as development that differs from normative, control-group development. This framework allows for interactions between treatment and initial status in their effects on development. Finally, an approach for the estimation of power to detect treatment effects in this framework is demonstrated. Illustrations of power calculations are carried out with artificial data, varying the sample sizes, number of timepoints, and treatment effect sizes. Real data are used to illustrate analysis strategies and power calculations. Further modeling extensions are discussed.

Bengt O. Muthén, Graduate School of Education & Information Studies, University of California, Los Angeles; Patrick J. Curran, Department of Psychology, Duke University.

This research was supported by grants from the National Institute of Mental Health (40859) from the Office of Educational Research and Improvement to the National Center for Research on Evaluation, Standards, and Student Testing, and from the National Institute of Alcohol Abuse and Addiction (K02 AA 00230-01 and F32 AA05402-10).

We thank Sheppard Kellam and Johns Hopkins Prevention Research Center for providing data for illustrations, and Hendricks Brown, Andrea Hussong, Siek-Toon Khoo, Linda Muthén, Jeff Shih, Booil Jo, and Noah Yang for helpful comments and assistance. Earlier versions of this paper were presented at the Workshop on a Scientific Structure for the Emerging Field of Prevention Research, cohosted by the Johns Hopkins University Prevention Research Center and the Prevention Research Branch, National Institute of Mental Health, December 5–6, 1994 and at the Fifth Biennial Conference on Community Research and Action, June 14, 1995.

Correspondence concerning this article should be addressed to Bengt O. Muthén, Graduate School of Education & Information Studies, University of California, Los Angeles, California 90024-1521. Electronic mail may be sent via Internet to bmuthen@ucla.edu.

Analysis of longitudinal data responds to the need in many research areas for describing individual differences in development over time. Recently, a host of methodological contributions have been made and the implications for applications of this new technology have not been fully spelled out. The broad aim of this article is to contribute to this process by discussing new types of models and the power of detecting various effects by such models.

Analysis of individual differences in longitudinal data draws on several different methodological traditions with their own specific analysis focus, terminology, and software. This complexity may have impeded dissemination of new research methods in this area. Three traditions appear especially important: biostatistics, education, and psychometrics. In the biostatistics tradition, the keywords include repeated measurement, random-effects analysis of variance (ANOVA), the mixed model, and random coefficient modeling. Key references include Rao (1958), Laird and Ware (1982), and Diggle, Liang, and Zeger (1994). Key software products include BMDP5V, SAS PROC MIXED, MIXED, and MIXOR. The education tradition started out relatively independent of

the biostatistics tradition but is now beginning to merge with it. Here, keywords include slopes-as-outcomes, multilevel modeling, and hierarchical linear modeling (HLM). Key references include Cronbach (1976), Burstein (1980), Goldstein (1987, 1995), Bock (1989), Bryk and Raudenbush (1992), and Longford (1993). Key software products include MLn, HLM, and VARCL. A largely independent tradition is found in psychometrics with keywords such as latent curve analysis and latent variable structural equation modeling. Key references include Tucker (1958), Meredith and Tisak (1984, 1990), and McArdle and Epstein (1987). Key software products include Amos, CALIS, EQS, LISCOMP, LISREL, MECOSA, and Mx.

In the biostatistics and education traditions, the individual differences in growth or decline over time are captured by random coefficients. Because these coefficients are unknown quantities that vary across individuals, the psychometric tradition views them as latent variables. Linkages between the traditions have been described in Browne and DuToit (1991), Muthén (1983, 1991, 1993), Rogosa (1988), Rogosa and Willett (1985), and Willet and Sayer (1994).

The currently available procedures for the random coefficient approach have both strengths and weaknesses. A strength is that they draw on statistical estimation procedures that have been thoroughly studied over many years. A weakness is that the modeling in these traditions has been largely limited to a single response variable that does not accommodate the general analysis needs of developmental theories (see, however, Raudenbush, Rowan, & Kang, 1991; Goldstein, 1995). The latent variable approach can essentially be characterized in the opposite way. Although the estimation procedures are currently not well developed for sufficiently general cases, the modeling framework has much more of the generality that is needed to answer researchers' questions.

In terms of generality, a comparison between currently available analysis procedures and software within the random coefficient approach and the latent variable approach may distinguish between two components of the analysis: the observed and latent variable data structure versus the model structure. It appears that the latent variable approach has an edge with respect to modeling flexibility and this aspect will be further investigated in this paper. At the same time, however, it would appear that the random coefficient approach is currently more flexible with respect to the observed and latent variable data structure. One example of this flexibility concerns randomly varying within-subject designs including unequal intervals of observation, varied within-person distributions of time-varying covariates having random effects, and data missing at random. Another example is the incorporation of clustered designs for persons. Although the latent variable literature includes treatments of missing data (see, e.g., Arminger & Sobel, 1990; Muthén, Kaplan, & Hollis, 1987) and growth modeling with clustered data (see Muthén, 1997), the analyses do not yet allow the flexibility of the random coefficient approach.

The latent variable framework is considered further in this article for reasons of model flexibility. When translating the random coefficient model into the latent variable framework, one finds that the standard random coefficient growth curve model corresponds to a very limited latent variable model. The latent variables are not introduced to represent latent variable constructs in the traditional psychometric sense of being measured by multiple indicators at a single timepoint. Instead, observations at multiple timepoints of the same outcome variable are used to determine latent variables that represent the shapes of the individual curves. Formally, the corresponding latent variable model is a confirmatory factor analysis model with unusually restrictive factor loading constraints. Once the random coefficient growth curve model has been put into the latent variable framework, many general forms of longitudinal analysis can be studied, including mediational variables influencing the developmental process, ultimate (distal) outcome variables influenced by the developmental process, multiple developmental processes for more than one outcome variable, sequential-cohort and treatment-control multiple-population studies, and longitudinal analysis for latent variable constructs in the traditional psychometric sense of factor analytic measurement models for multiple indicators. The latent variable framework also accommodates missing data (see, e.g., Arminger & Sobel, 1990; Muthén et al., 1987), categorical and other nonnormal variable outcomes (see, e.g., Muthén, 1984, 1996), and techniques for clustered (multilevel) data (Muthén, 1994, 1997; Muthén & Satorra, 1995), but these features will not be discussed here.

The full potential of the more general longitudinal modeling that can be carried out within the latent variable framework has not yet been realized in terms of real-data analyses of substantive research questions. One aim of this article is to speed up this pro-

cess by outlining some nonstandard, prototypical models within the latent variable framework. Emphasis will be placed on a model in an especially challenging area, the case of longitudinal modeling within a true experimental design. These types of designs are often encountered in behavioral research in the form of prevention studies or intervention programs in mental health or evaluations of educational programs. Intervention programs are often characterized by community-based participant recruitment (e.g., schools, courts, government agencies) and the dissemination of treatment is through a field setting (e.g., classrooms, after-school groups, etc.; see Brown, Kellam, & Liao, 1996, for an overview). Developmental studies with randomized interventions make particularly good use of the longitudinal research design and warrant further methodological attention.

It is expensive and time-consuming to carry out longitudinal studies and particularly so with a large number of participants. It is therefore important to know the minimum number of participants and time-points that can be used to answer the research questions. In planning a longitudinal study, it is critical to estimate the power to detect certain effects, such as treatment effects in intervention studies. Little is known, however, about power issues for longitudinal modeling in general and intervention effects in particular, especially in the more general settings outlined above. A second aim of this article is therefore to present some relevant power results for longitudinal modeling in intervention studies. This article uses the general latent variable framework to consider power estimation using a method developed for latent variable models by Satorra and Saris (1985).

Although the article focuses on multiple-population longitudinal studies as motivated by an intervention context, it should be pointed out that the proposed growth analysis and power estimation techniques are generally applicable to multiple population settings. Such settings may for example involve gender differences and differences among populations varying in their risk for problematic development. The power estimation approach is also of interest in single population settings involving questions of power to detect certain growth patterns.

Latent Variable Longitudinal Modeling

A Conventional Random Coefficient Model

Using a simple random coefficient growth curve model as a starting point, a translation into latent variable modeling will be made to show the key features of the latent variable longitudinal model. The generalizations of this model will then be shown both in terms of the formulas for a general latent variable model as well as in terms of path diagrams. Throughout this article, we simplify the discussion by focusing on continuous normal outcome variables. It is clear, however, that methodology for categorical and other nonnormal variables is both very much needed in practice and has seen recent methodological advances.

Consider growth for a single outcome variable y observed for individual i at time-point t as related to a time-varying covariate v_{it} and a time-invariant covariate w_i. For simplicity, only one time-varying and one time-invariant covariate is considered. The key idea is that each individual has his or her own growth trajectory. Growth will be expressed in terms of a random coefficient model described as a two-level model. Level 1 is written as

$$y_{it} = a_i + b_i x_{it} + c_{it} v_{it} + e_{it}, \tag{1}$$

where a_i is an intercept, x_{it} is a time-related variable (such as age or grade), b_i and c_{it} are slopes, v_{it} is a time-varying covariate, and e_{it} is a residual. Level 2 is written as

$$\begin{cases} a_i = a + d_a w_i + e_{ai} \\ b_i = b + d_b w_i + e_{bi} \\ c_{it} = c_t, \end{cases} \tag{2}$$

where a, b, d_a, and d_b are (fixed) intercept and slope parameters, w_i is a time-invariant covariate, and e_{ai} and e_{bi} are residuals. It is assumed that the e_{it} are uncorrelated with e_{ai} and e_{bi}, whereas the latter two residuals may be correlated with one another.

This random coefficient model may, for example, describe linear growth over time by using the x_{it} coefficients 0, 1, 2, . . . in Equation 1, so that a_i represents the initial status of individual i on his or her growth trajectory and b_i represents his or her linear growth rate on this trajectory. It is straight-forward to add nonlinear growth by adding terms that are nonlinear in x_{it}. Note, however, that we are not considering models that are nonlinear in the random coefficients (for such model, see, e.g., Browne & Du Toit, 1991). The time-invariant covariate w explains part of the variation in each individual's growth trajectory by Equation 2. Each individual's growth trajectory is also influenced by the time-varying covariate v as seen in Equation 1.

The specific random coefficient model that is con-

sidered here is a conventional model for growth curve modeling in biostatistics with two exceptions. First, as indicated above, $c_{it} = c_t$ so that the coefficients for the time-varying covariates are not allowed to vary across individuals. This restriction is necessary to fit the random coefficient model into the conventional latent variable modeling framework given that this framework cannot handle products of random variables such as $c_{it}v_{it}$. Second, for the same reason, we assume that $x_{it} = x_t$, which means that all individuals are observed at the same timepoints. It should be pointed out that the latter restriction is in principle not necessary in the latent variable framework when different individuals are observed at different timepoints because of missing data such that different individuals may have different numbers of observations. Muthén et al. (1987) discussed missing data techniques that can be used in conventional structural equation modeling software when there is a small number of missing data patterns and where each missing data pattern is represented by a sizable number of observations. Recent developments in structural equation modeling software also incorporate the contrary case, allowing for individually varying missing data patterns (see also Arminger & Sobel, 1990).

In the latent variable tradition, the random coefficients a_i and b_i in the model of Equations 1 and 2 are reconceptualized as latent variables, that is, factors. This idea was introduced as *latent curve analysis* by Meredith and Tisak (1984, 1990) and we use this general term from now on. The term has an advantage over the common term *growth modeling* in that it represents modeling of individual curves that correspond not only to monotonic growth but also to stability, decline, and combinations thereof. For a pedagogical description of latent curve analysis, see McArdle (1988), Willet and Sayer (1994), and Muthén (1995); this description will be only briefly restated here and readers new to the area are referred to those papers for more detail. It is convenient to view the model in terms of conventional latent variable modeling path diagrams. Figure 1 shows this particular model for five timepoints and a time-varying covariate v that only influences the last three timepoints. In this and subsequent latent curve figures, the subscript i denoting variation across individuals is suppressed for both manifest and latent variables but should be understood.

It is clear from Figure 1 that the latent curve model can be easily translated into input for existing latent variable modeling software. The c coefficients of Equation 1 are the slopes for the v covariate at the different timepoints and the d coefficients of Equation 2 are slopes for the w covariate. At the different timepoints, the outcome variable y_{it} is related to the two latent curve factors with slopes of unity for the a_i factor and x_t for the b_i factor. It should be noted that the x_t scores need not be predetermined, fixed values. The latent variable framework makes it clear that these values are slopes that can be estimated, allowing for a flexible model to represent nonlinear trajectories. For example, the scores 0, 1, x_3, x_4, x_5 may be used, where x_3, x_4, and x_5 are parameters to be estimated. Estimates of these parameters can be compared with the unit step from the first to the second timepoint.

Further details of the model in Figure 1 are as follows. The intercept (or mean if there is no time-invariant covariate) for the a_i factor is a and the intercept (mean) for the b_i factor is b. Allowing an intercept (mean) parameter a for the a_i factor, the intercepts in the regressions of each of the five y_t variables on the factors and the time-varying covariate should be held fixed at zero. An alternative that will be used in this paper is to set the a mean of the a_i factor at zero in Equation 2 and instead let there be a common intercept in the y_{it} regression Equation 1. The common intercept is obtained by restricting the intercepts for each of the five y_t regressions to be equal across time. This alternative parameterization is equivalent to the original, but it generalizes in a more straightforward fashion to multiple-indicator models and models with interactions between treatment and initial status.

Further parameters are the two residual variances and the residual covariance for the two factors (these parameters constitute the factor covariance matrix elements if there is no time-invariant covariate). The residuals in Equation 1 may be allowed to have different variances across time and to have certain patterns of correlation across time. They are uncorrelated with the residuals in Equation 2. Identification issues are discussed in Muthén (1995). The model imposes a structure on both the mean vector and covariance matrix for the observed variables. The model specification is discussed further below within a general framework. Input for the LISCOMP program (Muthén, 1984, 1987, 1989) used in this article is given in Part 1 of a technical appendix, which is available from either Bengt O. Muthén or Patrick J. Curran. The same results would be obtained with Amos, EQS, and LISREL.

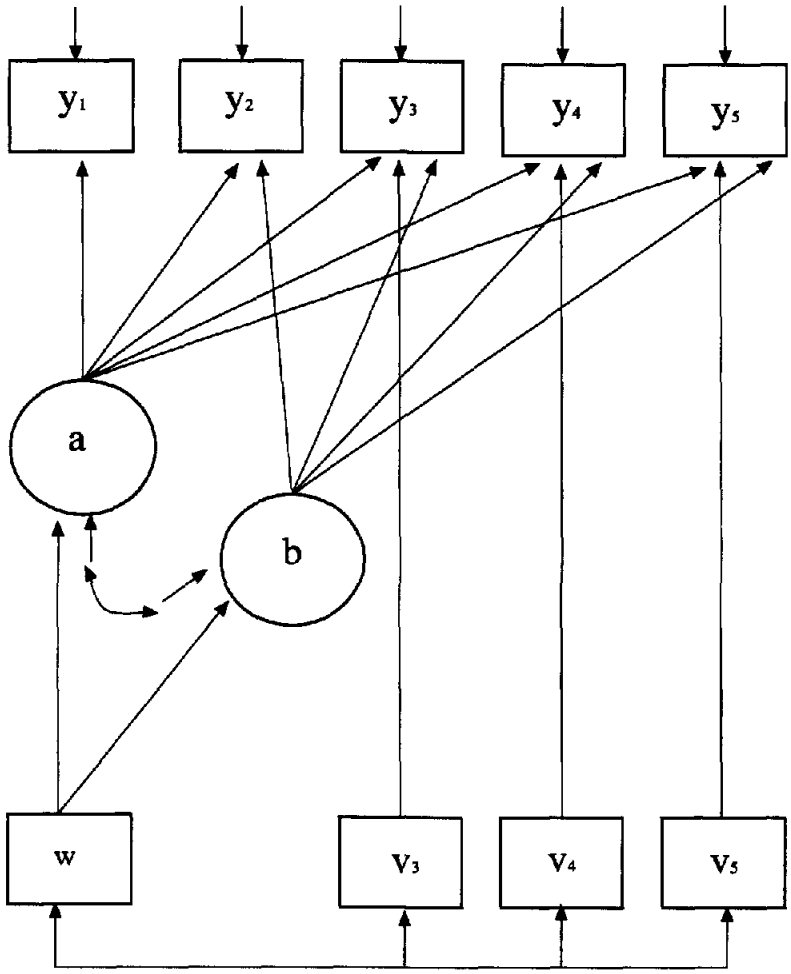

Figure 1. Latent variable growth model for five timepoints with a time-invariant and a time-varying covariate.

A General Latent Variable Framework

The latent curve model described above fits into the following general latent variable framework (cf. Bollen, 1989; Jöreskog & Sörbom, 1979). For population (group) g, consider a p-dimensional observed variable vector y^g related to an m-dimensional latent variable vector η^g through a factor-analytic measurement model,

$$y^g = \nu^g + \Lambda^g \eta^g + \epsilon^g, \tag{3}$$

where ν^g is a vector of measurement intercepts, Λ^g is a $p \times m$-dimensional matrix of measurement slopes (factor loadings), and ϵ^g is a p-dimensional vector of measurement residuals. Here, $V(\epsilon^g) = \Theta^g$. The latent variables have the structural relations

$$\eta^g = \alpha^g + B^g \eta^g + \zeta^g, \tag{4}$$

where α^g is an m-dimensional vector of structural intercepts (for endogenous ηs) or means (for exogenous ηs), B^g is an m-dimensional matrix of structural slopes, and ζ^g is an m-dimensional vector of structural residuals. Here, $V(\zeta^g) = \Psi^g$, a residual (for endogenous ηs) or latent variable (for exogenous ηs) covariance matrix.

Under regular assumptions on the residuals, we have the mean and covariance structure

$$E(y^g) = \mu^g = \nu^g + \Lambda^g (I - B^g)^{-1} \alpha^g \tag{5}$$

and

$$V(y^g) = \Sigma^g = \Lambda^g (I - B^g)^{-1} \Psi^g (I - B^g)^{-1'} \Lambda^{g'} + \Theta^g. \tag{6}$$

This is the standard multiple-population structural equation modeling formulation used in LISREL-type

modeling. With the customary assumption of i.i.d. sampling from each of the G populations, a simultaneous, multiple-population analysis is commonly achieved by minimizing the fitting function F,

$$F = \sum_{g=1}^{G} \{N^g[\ln |\Sigma^g| + tr(\Sigma^{g^{-1}} T^g) - \ln |S^g| - p]\}/(N-1), \quad (7)$$

where N is the total sample size and

$$T^g = S^g + (\bar{y}^g - \mu^g)(\bar{y}^g - \mu^g)', \quad (8)$$

which gives maximum-likelihood estimation under multivariate normality for y^g (see, e.g., Jöreskog & Sörbom, 1979; Sörbom, 1982). At the optimal value of F, $(N-1)F$ is asymptotically distributed as a chi-square variable.

The simple example in Figure 1 of the previous subsection may be fitted into this modeling framework by letting y in Equation 3 contain all nine observed variables and η contain the two curve factors as well as one factor corresponding to each observed variable, where the latter factors are measured without error. This specification is explicated in Part 1 of the technical appendix, which can be obtained from Bengt O. Muthén or Patrick J. Curran. Note that this specification prepares for multiple indicators of latent variable constructs, both among the outcome variables and among the covariates. In the Figure 1 example, the parameters of Equations 5 and 6 are as follows: $\nu = 0$, Λ contains 0s and 1s, $\Theta = 0$, B contains the coefficients $(1x_t)$, α contains the equal intercepts in the regressions of each dependent-variable ηs and the means of the covariates, and Ψ contains the residual covariance matrix for the dependent-variable ηs and the covariance matrix for the covariates. For an introduction to this specification, see Willet and Sayer (1994) and Muthén (1995).

Some Nonstandard, Prototypical Models

Following are some potential applications of the general latent variable framework given above that generalize the conventional random coefficient growth model discussed in an earlier section. These are given as examples of model types that are likely to play important roles in future longitudinal analyses. Each of these examples can be further extended to include latent variable constructs η measured by multiple indicators as in Equation 3, where these constructs assume the role of outcome variables, covariates, mediators, and ultimate outcomes. All constitute examples that can be fit into existing software technology and illustrate that currently, the latent variable approach to growth modeling appears to be more flexible than random coefficient modeling techniques.

Figure 2 shows a latent curve model where the influence of the time-invariant covariate w is mediated by a variable z. It is of interest to study to which extent the influence of w on the two curve factors is indirect through z and to which extent there is a direct influence. This has practical importance when the mediator is a variable that can be manipulated and the covariate cannot be manipulated. In this case, the extent to which there is an indirect influence through the mediator indicates the extent to which an intervention aimed at manipulating the mediator can have an influence on the developmental process.

Figure 3 shows a latent curve model where an ultimate outcome variable y_6 is influenced by the two curve factors of a previously observed outcome variable. In this model, the trajectory of an individual, not his or her scores on the outcome variable, is the predictor of the ultimate outcome variable.

Figure 4 shows a latent curve model that may be seen as a generalization of that in Figure 3. Here, one developmental process preceeds and predicts the course of a second, later occurring, developmental process. Muthén (1997), Stoolmiller (1994), Curran, Harford, and Muthén (1996), and Curran, Stice, and Chassin (1997) used variations of this model where the two developmental processes were concurrent and where the initial status factor of each process was hypothesized to influence the growth rate factor of the other process.

A Nonstandard, Prototypical Model for Intervention Studies

The types of latent curve models shown in Figures 1–4 can all be generalized to the simultaneous analysis of data from several populations (i.e., multiple-population analysis). To a limited extent, population differences can be captured in single-population analysis by representing the groups as dummy variables used as time-invariant covariates. Although this enables population differences in location for the curve factors, other parameters are not allowed population differences (we note in passing that representing groups by time-varying dummy variables does not describe population differences among individuals but changes in group membership for the same individual). To achieve more generality in the modeling,

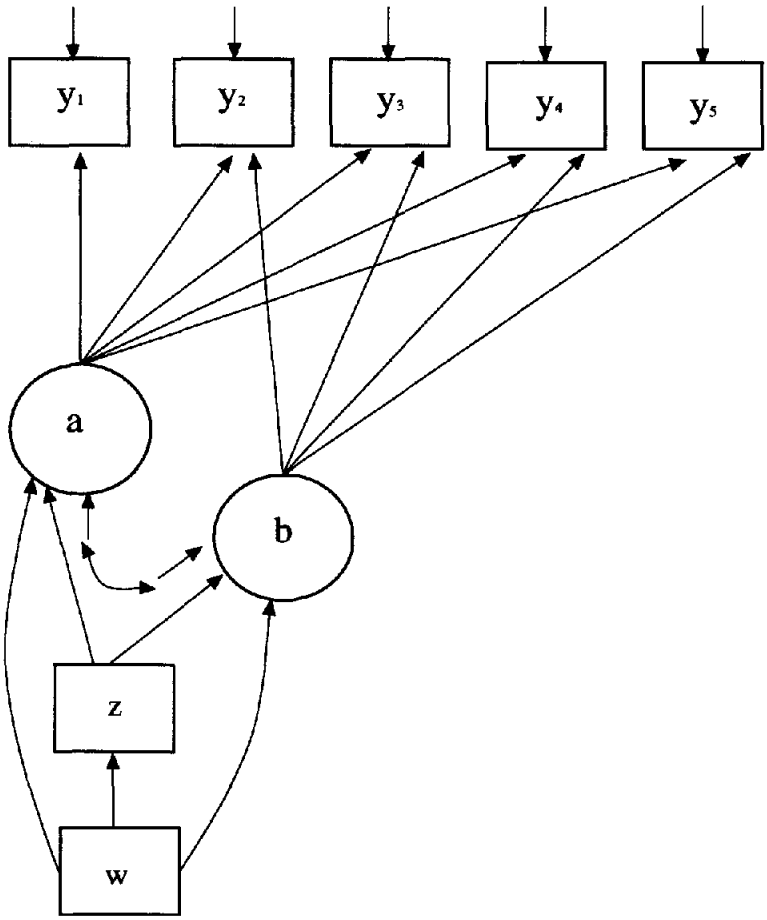

Figure 2. Growth model with a mediating variable.

however, researchers need to use a multiple-population approach instead of a dummy-variable approach. This is particularly beneficial in the setting of intervention studies.

A Two-Group Formulation

Consider an intervention study where individuals are measured before being randomized into a treatment or a control group and then measured repeatedly thereafter. In line with Jöreskog and Sörbom (1979), this may be viewed as data from two different populations. The control group population represents the normative set of individual trajectories that would have been observed also in the treatment group had they not been chosen for treatment. The effect of treatment is assessed by comparing the set of trajectories in the treatment population with those in the control population.

This two-group setting may be described in path diagram form as shown in Figure 5. This is readily generalized to the case where there are several treatment groups. In Figure 5, the top graph represents the control group, where for simplicity we may assume linear growth, using an initial status factor and a linear growth rate factor. Also for simplicity, no covariates are included.

The bottom graph in Figure 5 represents the treatment group. In line with conventional multiple-population latent variable analysis, we could specify a two-factor growth model also here and test for equality of parameters across the two populations. Lack of equality would then be taken as evidence of effects of treatment. There is, however, a better alternative which offers a more useful analysis with respect to treatment effects. This alternative is shown in Figure 5. Here, an additional growth factor is introduced for the treatment population. Whereas the first two factors are the same as for the control group, the third

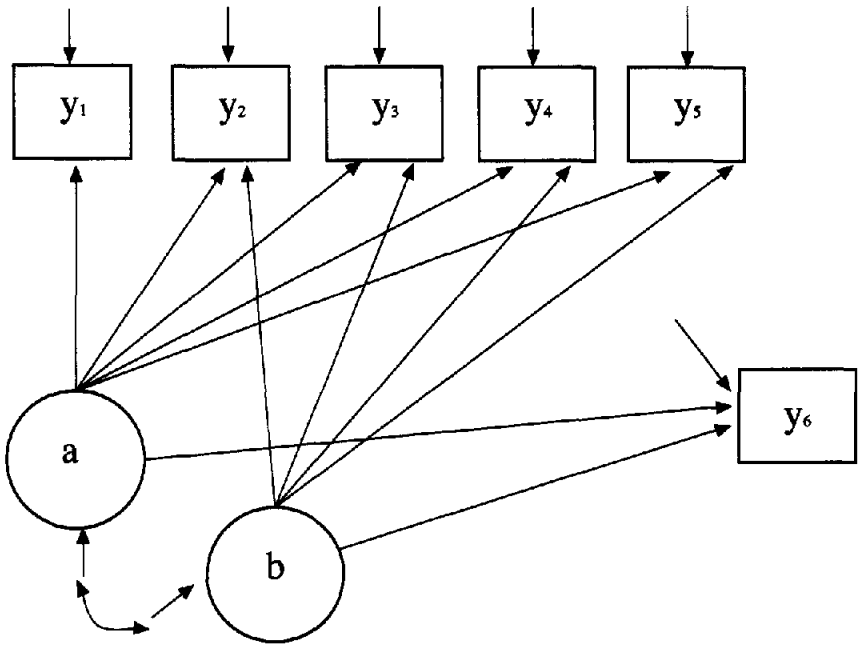

Figure 3. Growth model with a distal outcome variable.

factor represents incremental or decremental growth or decline that is specific to the treatment group. We will call this factor the added curve factor due to treatment.

The interpretation of the three factors in the treatment group can be described as follows. For simplicity, the intervention is assumed to take place after the initial timepoint. Using the x_t scoring of 0, 1, 2, 3, and 4, only the first factor influences y_1. At subsequent time-points, the second factor gives additional contributions. The first factor can therefore be interpreted as the individual's initial status before the intervention started, whereas the second factor represents the growth rate. This is the growth rate that the individual would progress according to had he or she not received the treatment. The parameters for the first and second growth factors in the treatment population are constrained to be equal to those of the control population. Thereby, the added growth factor captures the incremental or decremental growth beyond that of the control population. The treatment effect is thereby expressed in this added third growth factor that is specific to the treatment population. The treatment effect can be characterized by the mean of the added growth factor, adding or subtracting to the control growth rate. It can, however, also be characterized by the variance of the added growth factor, where an increased variance represents a treatment effect that makes growth more heterogeneous among individuals. Note that a treatment effect inducing a smaller variance in the outcome variable can also be represented in this model. This is achieved by a negative covariance between the first and the third growth factors.

Restricting the parameters of the first two growth factors to be equal across the control and treatment populations is warranted in a randomized intervention study. Often, however, the randomization breaks down during the course of the study and the two groups do differ significantly at the pre-intervention timepoint (cf. Cook & Campbell, 1979). In such cases, the equality constraints related to the initial status factor should be relaxed. First of all, this involves relaxing the mean of zero for the initial status factor in the treatment population. The intercepts in the y regressions are still held equal over time and across populations, but the treatment initial status mean is thereby allowed to be different from the zero value of the control population. Second, this involves relaxing the across-population equality constraint for the initial status variance. The above approach can still be used as long as it is realistic to assume invariance across the two populations for the parameters of the second growth factor. As usual in nonexperimental studies, the realism of the modeling can be improved by using time-invariant covariates so that the

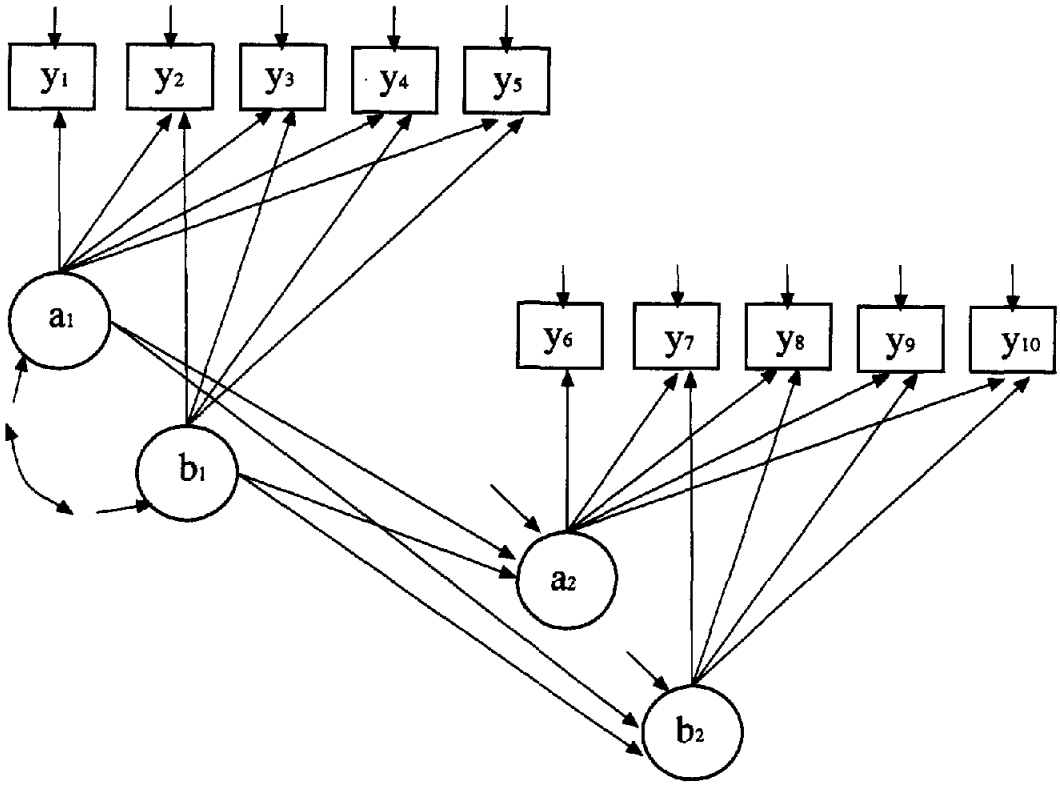

Figure 4. One growth process influencing a later growth process.

equality constraints are instead applied to parameters describing the conditional means and (co-)variances of the two growth factors.

The effect of treatment may be more complex than merely changing the growth rate of a process that has the same functional form as that of the control group (i.e., a line in the above example). For example, the control group may follow linear trajectories whereas the treatment group may follow nonlinear trajectories. The third factor may in this case be represented for example by a quadratic growth term, using x_t^2 scoring of 0, 1, 4, 9, and 16. Another solution, offering more flexibility, is to use estimated x_t scores as discussed earlier.

The above discussion focuses on treatment effects that are permanent in the sense that the differences between the two average trajectories keep increasing over time. It is also important to be able to capture temporary treatment effects, given that such effects are probably more common in intervention studies. The approach of estimating x_t scores is useful here given that these scores are allowed to first increase and then decrease. If there is a specific hypothesis for when a treatment effect begins or ends, one can instead use piece-wise curve factors that influence the outcomes only at certain timepoints (for piece-wise linear modeling, see, e.g., Bryk & Raudenbush, 1992; Seltzer, Frank & Bryk, 1994).

Treatment-Initial Status Interactions

We describe an extension of the above two-group latent curve model that responds to a central concern of intervention studies, namely trying to understand for whom an intervention is effective. Our formulation is related in spirit to both the Bryk and Weisberg (1976) valued-added analysis and the Rogosa (1991) discussion of how to view treatment interactions in the context of aptitude-treatment interactions. It is frequently the case that individuals at different pre-intervention (baseline) levels on the outcome variable benefit differently from the intervention (see, e.g., Cronbach & Snow, 1977). In analysis of covariance (ANCOVA) studies using pre- and postintervention measures, this is often studied in terms of an interaction between the baseline and the treatment, using the baseline as a covariate. In longitudinal modeling, the initial status factor provides a more relevant covariate. The baseline variable may in fact be seen as a fallible

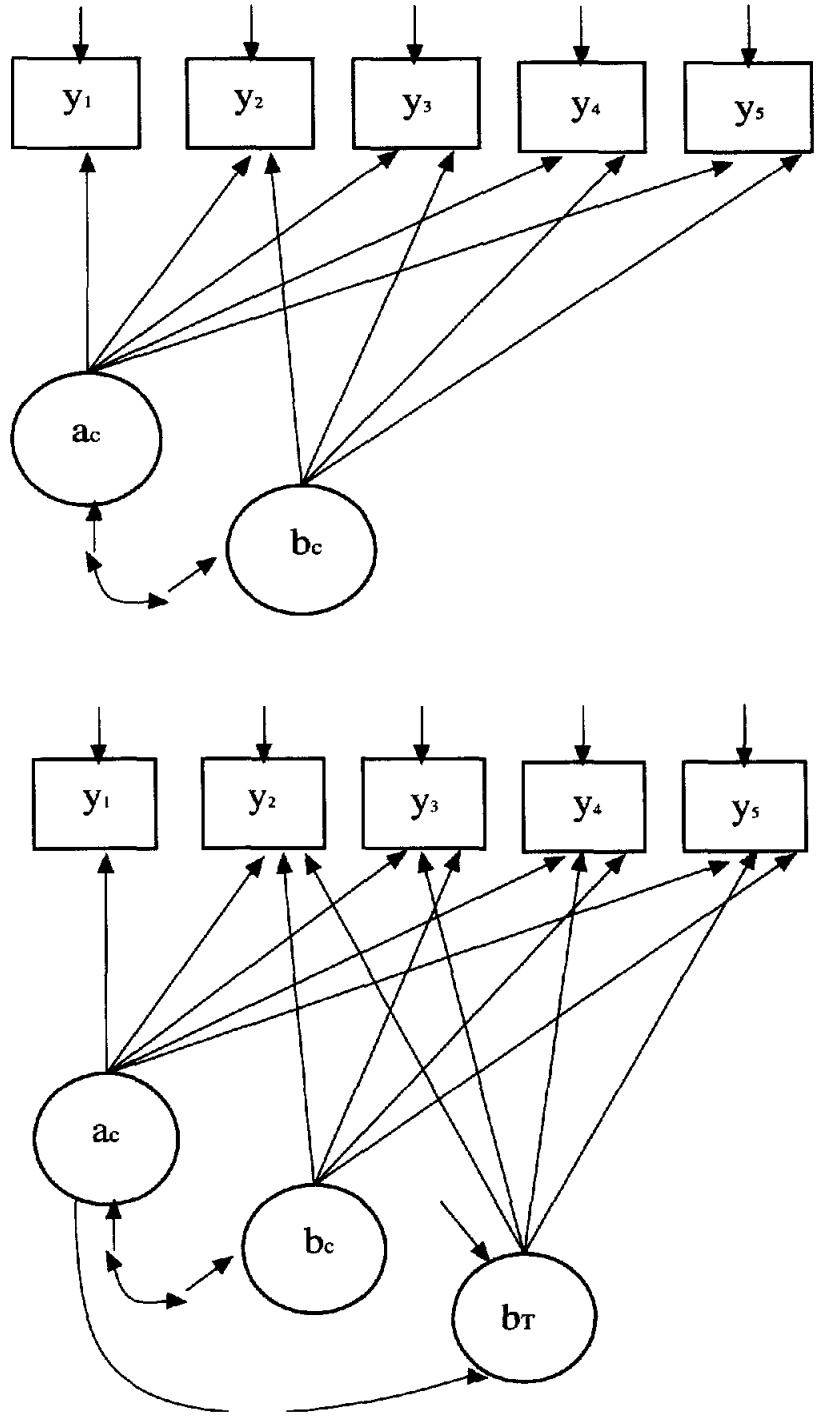

Figure 5. A two-group growth model for intervention studies.

indicator of this factor. These concerns lead to longitudinal modeling of interactions involving the treatment and the initial status in their influence on the rate of change.

In terms of the Figure 5 latent curve model, the interaction may be expressed by letting the initial status factor influence the added growth factor in the treatment population. For example, in a remedial

reading program, a negative influence may be viewed as lower initial status individuals having larger incremental growth rate effects from the treatment. In the latent variable framework, the influence is expressed in terms of a structural regression with the added growth factor as the dependent variable and the initial status factor as the independent variable. This is a logical formulation given that the initial status factor has temporal precedence over the added growth factor. Including this structural regression in the model, the treatment effect can be described in an even richer way. The mean and variance of the added growth factor is then expressed as a function of the initial status mean and variance, the structural regression intercept and slope, and the residual variance in that regression.

It may seem paradoxical that an interaction can be described by a structural regression that is a linear function of initial status, but it should be kept in mind that the regression is formulated within a simultaneous analysis of the control and treatment groups, where the control group does not include this regression, thereby inducing the interaction. The multiple-group approach of linear structural equation modeling thus enables interaction modeling for a continuous latent variable when the other variable involved is an observed categorical grouping variable as with the treatment–control dichotomy. Interactions among continuous latent variables, however, require special techniques (see, e.g., Jaccard & Wan, 1996).

The intercept and slope in the added structural regression have clear interpretations in terms of main effect and interaction effect of treatment. If the slope is zero, the intercept represents the mean of the added growth factor given that the initial status factor mean is zero. The intercept therefore represents the main effect of treatment. With a nonzero slope, the conditional mean of the added growth factor given the initial status factor at its mean value of zero is still equal to the intercept (the main effect). The slope value indicates to what extent changes away from the mean of the initial status factor influences the added growth factor beyond its main effect (intercept) value. The slope therefore represents the interaction effect that the treatment induces.

It may be noted that a negative slope value represents a treatment that produces a more homogeneous outcome. The negative slope serves to reduce the variance of the outcome variable in the treatment group because individuals with high initial values tend to get lower growth rates and individuals with low initial values tend to get higher growth rates.

It is also possible to include a more complex interaction using this two-group framework. In the treatment population we may allow not only the first but also the second growth factor to influence the added growth. The second factor represents the normative growth rate the individual would have had without treatment. For example, if the normative growth rate is low for the individual, the added treatment growth rate may be high, representing a negative influence from the second factor to the added growth factor.

It is worth pointing out two methodological issues about this approach to treatment interactions. First, it is made possible by the specification of an added growth factor specific to the treatment population so that the treatment effect is separated from normative growth. Second, the approach of regressions among growth factors is currently unique to the latent variable approach. For example, it is not possible in the conventional random coefficient specification of Equations 1 and 2 to represent a treatment initial status interaction by including a_i on the right-hand side of the equation for b_i.

Analysis Strategies

Given the complexity of the proposed approach to intervention analysis, a careful analysis strategy is required. Five analysis steps are discussed here.

As a first step, the normative development can be studied by a separate analysis of the control group. Previous research may have established a priori hypotheses about the form of these trajectories. Inspection of individual and overall developmental patterns may also contribute to choosing the models attempted in the analyses. In this single-group analysis, it is valuable to rule out that the control population exhibits any of the postintervention changes in trajectories that are hypothesized to be due to treatment. In this way, if such trajectory changes are found for the treatment population, they are more clearly attributable to the treatment.

As a second step, the treatment group can be analyzed separately. Here, the basic trajectory form (linear, nonlinear) may be investigated. The treatment may induce curve shapes different from those in the control group.

As a third step, a two-group analysis is performed where the latent curve factors found for the control group are repeated in the treatment group. For all control factors but the initial status factor, one may specify an added treatment factor. For example, the control population may have both a linear and a qua-

dratic growth factor beyond the initial status factor. In this case, the treatment population may have added factors for both linear and quadratic growth, or the control population may have only linear growth, in which case it may suffice to have an added quadratic factor in the treatment population.

As a fourth step, treatment interaction is tested for in the two-group analysis. Here, the initial status factor in the treatment group is allowed to influence the treatment group growth factors.

As a fifth and final step, a sensitivity analysis is carried out for the two-group model from Step 4. First, the model is scrutinized in terms of successful randomization. Deviations from perfect randomization can be allowed for by relaxing the equality restrictions with respect to the initial status factor. Second, tests of overall treatment effect are carried out with treatment effect parameters set to zero. The assessment of treatment effects can thereby be expressed as a chi-square difference test for the models of Steps 3 and 4. There are several treatment effect parameters: interactions, main effects, and variance effects. For example, if an interaction is captured by the initial status influencing the added linear growth factor, the interaction parameter is the slope in this structural regression, the main effect is the intercept, and the residual variance represents the variance effect. Using this example, the testing may be done in stages, first restricting the slope to zero, and second, if this is found tenable, restricting the intercept to zero.

Power Estimation in the Latent Variable Framework

The estimation of power to detect misspecified latent variable models has been discussed in Satorra and Saris (1985) and Saris and Satorra (1993); see also Saris and Stronkhorst (1984). In principle, power can be estimated for any model by carrying out a Monte Carlo study that records the proportion of replications in which the incorrect model is rejected. Satorra and Saris proposed a method that gives a tremendous simplification over such a brute force approach. A key technique is based on the likelihood-ratio chi-square test for maximum-likelihood estimation of mean and covariance structure models such as the one in Equations 5 and 6. Here, we will use this technique to estimate the power to detect intervention effects in the two-group latent curve model discussed above. The Satorra–Saris approach is particularly suitable for the intervention setting given that power estimates are desired for very specific model misspecifications concerning the absence of treatment effects. MacCallum, Browne, and Sugawara (1996) discussed power estimation techniques that concern overall model fit, but that will not be considered here.

Under multivariate normality for y^g, $(N-1)F_{min}$, where F_{min} is the optimal value in Equation 7, is distributed asymptotically as a chi-square variate when the model in Equations 5 and 6 is correct. Satorra and Saris (1995) showed that when the model is incorrect but not highly misspecified, $(N-1)F_{min}$ is asymptotically distributed as a noncentral chi-square variate with a certain noncentrality parameter, which can be approximated by a two-step procedure. This procedure involves two models, one more general that is assumed correctly specified and one more restrictive that is misspecified.

In our intervention setting, we are interested in the power to detect intervention effects and the more restrictive model sets the corresponding parameter(s) to zero. As a first step, the more general two-group latent curve model is estimated including the treatment effect(s). In a second step, the estimated mean vectors and covariance matrices from Step 1 are used in place of the corresponding sample statistics and analyzed by the more restrictive model that sets the treatment effect parameter(s) to zero. The value of $(N-1)F_{min}$ in this second step represents an approximation to the noncentrality parameter. Once this parameter has been obtained, the power can be obtained from tables for noncentral chi-square distributions as a function of the degrees of freedom and the α level of the test (see, e.g., Saris & Stronkhorst, 1984). A technical appendix, which may be obtained from Bengt O. Muthén or Patrick J. Curran, gives a short SAS IML program that computes the power in this way. The degrees of freedom refer to the number of treatment effect parameters.

Saris and Satorra (1993) point to simulation studies that indicate that this procedure for estimating power can be sufficiently accurate for practical purposes at small sample sizes. A simulation study by Curran (1994) found very good results at sample sizes of 100. To verify the accuracy for the present two-group growth model, we carried out a limited simulation study. Data were generated over 1,000 replications from the Figure 5 linear growth model with a certain sample size and treatment main effect size (defined below). The proportion of the replications for which the t value of the treatment effect exceeded its 5% critical value was recorded. This t value refers to the incorrect hypothesis of zero treatment effect. The

study focused on power values close to 0.80, varying the sample size. For a treatment effect size of 0.30 and a total sample of 200 divided equally among control and treatment group observations, the Satorra-Saris method obtained a power of 0.734 (this is the result shown below in Figure 9, curve C) as compared with 0.755 from the simulation. An even better agreement was obtained at the higher total sample size of 500 with a treatment effect size of 0.20 where the Satorra–Saris method obtained a power of 0.783 (see Figure 6, curve B), whereas the simulation resulted in 0.780.

In this article, the Satorra–Saris method for estimating power will be used to compute power curves as a function of sample size for a variety of hypothetical two-group models of the type shown in Figure 5. Here, parameter values will be chosen to represent various treatment effect sizes. These values generate the mean vectors and covariance matrices that are used in the second step of the power method. The method will also be used in connection with the real-data analysis. Here, the parameter estimates obtained from an analysis of the real data are taken to represent population values that the analyst believes are meaningful for power analysis. These values are used to generate the mean vectors and covariance matrices for the second step of the power method.

Analysis of Examples

The general analytic and power estimation framework will now be illustrated. First, power curves will

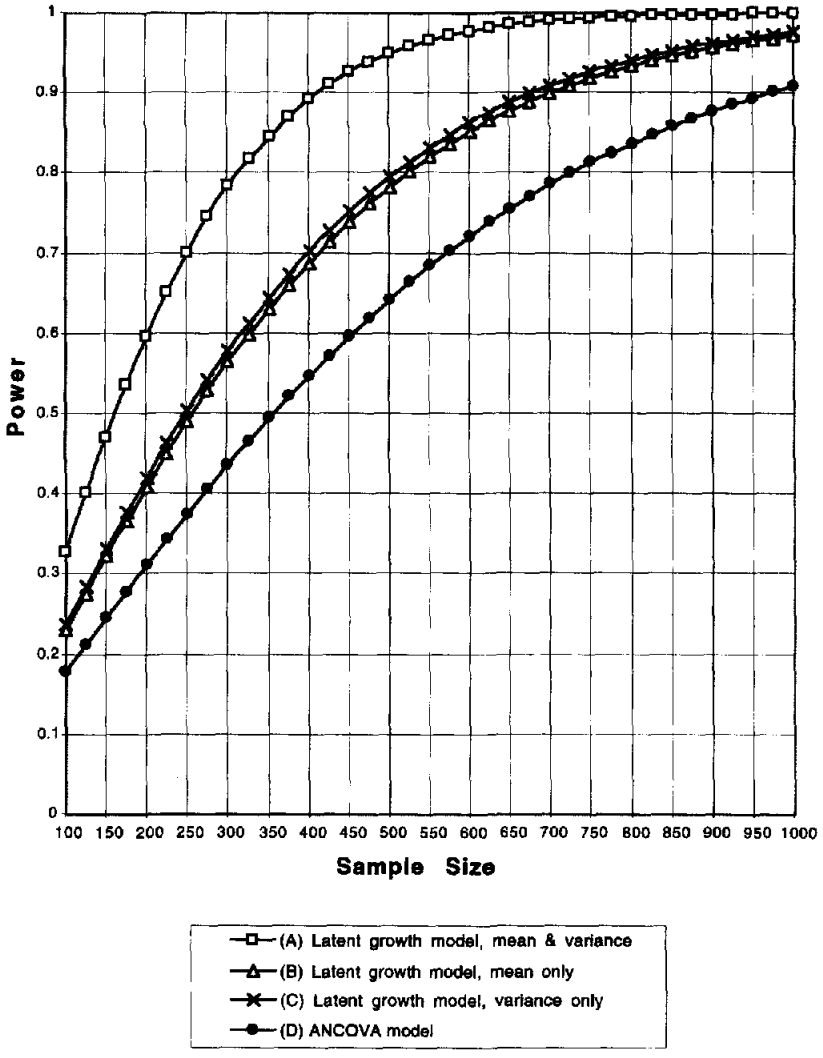

Figure 6. Power to detect a main effect of $ES = .20$ assessed at Time 5.

be calculated for various artificial models, including models with an interaction between treatment and initial status. Second, a real-data example with this type of interaction is analyzed and the power to detect the intervention effects is estimated.

Artificial Data: Power Curves

In this section, a set of power curves will be shown for different cases of the two-group intervention linear growth model of Figure 5. Many different situations are in principle of interest: We may have an experimental study with individuals randomized into treatment and control groups, or the study may be nonexperimental with pre-existing differences measured by covariates; the study may have a balanced design or not; the treatment effect may be permanent or temporary; and there may be interactions between the treatment and the initial status, or not. To limit space, we did not consider nonexperimental studies with covariates and temporary treatment effects.

The calculation of power curves calls for a consideration of effect size (Cohen, 1988). In a traditional two-group t-test setting, effect size is typically defined as the treatment and control group difference in outcome means, divided by a standard deviation based on the pooled outcome variance. A small effect size is typically taken to be 0.20, a medium effect size 0.50, and a large effect size 0.80 (Cohen, 1988). In the latent curve model setting, the definition of effect size is not as straightforward. First, although Cohen-type definitions concern manifest variables, treatment effects in our models can also be expressed in terms of latent variables. For example, in the Figure 5 model, the treatment effect may be expressed in terms of the mean difference for y at the last timepoint or in terms of the increase in the growth factor mean due to the added growth factor in the treatment group. Second, if reporting Cohen-like effect sizes for manifest variables, the standard deviation could be based on the control group rather than pooling over the treatment and control groups. The control group provides the normative value, whereas the treatment group variance in part reflects the treatment effect. In this article, we report effect sizes in several of these metrics.

The power calculations to be illustrated below raise the issue of how low the sample size can be for trustworthy analysis results given the dependence on asymptotic theory. Here, it should be noted that considerations of power may suggest sample sizes that are smaller than what can be recommended for obtaining good estimates of parameters and standard errors. For example, the simple artificial growth model of Figure 5 has 10 parameters in the control group. A conventional requirement in the latent variable literature is 5–10 observations per parameter (see, e.g., Bentler & Chou, 1988). Interpreting this as 5–10 individuals per parameter is probably too strong of a requirement given that each individual contributes serveral observations over time, which, although not independent, are not correlated 1.0. Using this rule of thumb, however, would lead to a minimum of 50 and preferrably 100 control group observations. Although the treatment group contributes information to the estimation of parameters describing the normative development in the two-group analysis, the risk of model misspecifications can be reduced by determining normative development from control group observations alone. With a balanced design, a total of 100–200 control and treatment group observations may therefore be desired for this particular growth model. This total sample size requirement may exceed the number required for a power of at least 0.80 and this should be kept in mind when studying the power figures below.

No Treatment Interactions

Consider the two-group intervention latent curve model of Figure 5. The means and variances are the same in the control and treatment group at the first timepoint due to randomization and there is linear growth in both groups. In line with what is commonly seen in practice, the control group variance of the growth rate is set at 20% of the variance of its initial status factor. For the control group growth, the parameter values are chosen so that the growth over the five timepoints corresponds to one standard deviation at the fifth timepoint. The treatment group increase over the five timepoints is chosen to produce various effect sizes in the sense of Cohen (1988). The effect size is calculated here as the difference in treatment and control group means for y at the fifth timepoint divided by the square root of the variance at the fifth timepoint pooled across the control and treatment groups. This treatment effect is achieved by a nonzero mean and variance for the added growth factor (the third factor) in the treatment population. The added growth rate variance is taken to be the same as the control group growth rate variance. For a small effect size of 0.20 (Cohen, 1988), this means that in standard deviation units of the control group growth factor, the treatment group growth rate mean corresponds to an increase of .23 over the control group growth rate mean. The residuals are specified to have equal vari-

ances across time and to be uncorrelated across time. The correlations among the repeated measures across time range from .32 to .50 for the control group. The parameter values, mean vectors, and covariance matrices for this case are given in a technical appendix that may be obtained from Bengt O. Muthén or Patrick J. Curran.

In this setting, power may refer to the detection of different parts of the treatment–control population differences: (a) the detection of a growth rate mean difference; (b) the detection of a growth rate mean, variance difference, or both; and (c) the detection of a growth rate mean, variance difference, or growth residual variance differences. Using the Satorra–Saris (1985) method, power is estimated for the three cases by using the misspecified model that restricts the corresponding parameters to be equal across the two populations. For (a) and (b), this equality constraint amounts to fixing the mean and variance to zero for the third added growth factor in the treatment population. In most intervention studies, it is probably of central interest to focus only on the growth rate mean effect as in (a). This says that an intervention is successful only if the mean of the growth rate changes and not successful if only the variance of the growth rate changes. In (b), the variance of the growth rate is an additional concern and in some studies this may be the only treatment effect. Although for a given sample size, it will be seen that the power is considerably larger for (b) than for (a), it would seem that the design should strive for a sample size that gives sufficient power already for (a). The added concern of residual variances in (c) is probably of little interest in most studies because seldom can randomization be expected to work out well enough for residual variance differences to be attributed to treatment effects.

Treatment effects and ANCOVA. Figure 6 gives power curves for the five-timepoint linear growth model with effect size 0.20 and sample sizes ranging from 100 to 1,000. Here, sample size refers to the total number of individuals in the control and treatment group, divided equally (balanced case).

The top curve of Figure 6, curve A, corresponds to the power of detecting both a growth rate mean and variance effect, whereas curves B and C correspond to the power to detect a growth rate mean effect only and a growth rate variance effect only, respectively. It is seen that curves B and C do not differ greatly, whereas curve A shows considerably larger power. In this article, we focus on curves of type B concerning mean growth rate. For this linear growth model, curve B shows that a rather large total sample size of about 525 is needed to achieve a power of 0.80 for this small effect size. Figure 8 shows the corresponding curves for larger effect sizes.

For comparison, Figure 6 also gives the power curve, curve D, for detecting a treatment effect using conventional ANCOVA. Here, the y measurement at the last timepoint is the outcome variable and the y at the first timepoint is the covariate. The Satorra–Saris (1985) method in a two-group setting is used also here with the standard ANCOVA specification of equal slopes across groups for the covariate. Analogous to the growth model, the means and variances of the covariate are also held equal across groups while the residual variances are allowed to be different. The ANCOVA results are in this case very close to what would be obtained from a t test given that the correlation between the outcome variable and the covariate is only 0.28. Comparison of curve D with curve B shows the latent curve modeling advantage of using information from all timepoints versus using only the first and last timepoint as in ANCOVA. The sample size needed to achieve a power of 0.80 is about 725 for the ANCOVA model versus only 525 for the growth model.

From a design point of view, power curves such as these can be used for cost considerations. For example, the increased cost of the study due to a need for a larger sample size with ANCOVA than with growth modeling for a given power level can be weighed against the decreased cost due to needing only two measurement occasions with ANCOVA. Although using a more costly design, it should be noted that apart from the power advantage, growth modeling has distinct analysis advantages over ANCOVA in that the latter cannot capture the form of the growth from the first to the last timepoint nor discover any limitations in duration of treatment effects.

Study length and number of measurement occasions. The comparison of longitudinal modeling with ANCOVA raises the issue of how the number of timepoints affects power. There are three key aspects of this: the length of the study, the number of measurement occasions for a given study length, and the study length for a given number of measurement occasions. These three aspects will be illustrated in turn.

In Figure 7, the length of the study is varied as three, four, five, and seven timepoints. The figure only considers the power to detect the growth rate mean difference, corresponding to curve B in Figure 6. As in Figure 6, a small effect size of .20 at the fifth

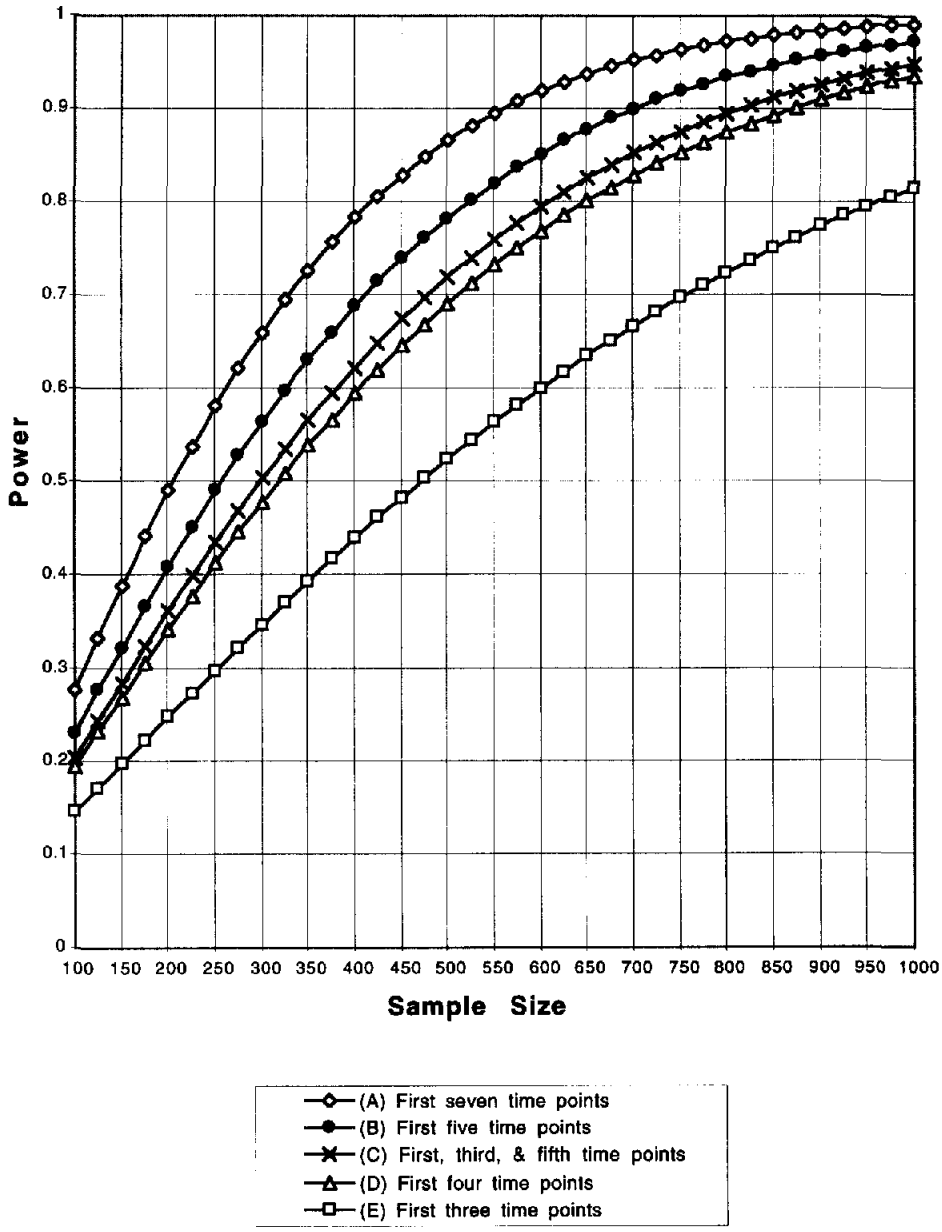

Figure 7. Power to detect a main effect of $ES = .20$ assessed at Time 5 varying as a function of total number of measurement occasions.

timepoint is considered. It is seen that the required sample size for obtaining a power of 0.80 drops sharply when extending the study length from three to four timepoints, with smaller drops when extending to five or seven timepoints. Given that effect size is defined at Timepoint 5, study length and effect size are to some extent confounded because effect size grows over time. The effect size does not, however, vary greatly over time because the variances increase over time. With a small effect size of .20 at Timepoint 5, effect sizes at Timepoints 3, 4, 6, and 7 are .159, .186, .209, and .214, respectively.

Figure 7 also illustrates the effect of number of measurement occasions for a given study length. Power curve C corresponds to using only three measurement occasions over the five timepoints, skipping Timepoints 2 and 4. This curve may be compared with that of the five-timepoint model for the same

study length (curve B). Here, there is no confounding in the sense that the two curves have the same effect size at Timepoint 5. The gain in using five versus three measurements is measured by the drop from 600 to 525 observations when requiring a power of .80. As noted above, ANCOVA at Timepoint 5 requires a sample of 725 for a power of .80. For a given study length, this shows that the largest gain is obtained by moving from the two-timepoint ANCOVA to the three-timepoint longitudinal modeling, whereas using more measurement occasions in the modeling gives diminishing returns. There are, however, clear advantages to using more than three timepoints in longitudinal modeling from the perspective of distinguishing between alternative growth forms (see, e.g., Muthén, 1995).

Figure 7 also shows the effect of study length for a given number of measurement occasions. Curve C and curve E both correspond to a three-timepoint growth model. Whereas curve C has measurements at Timepoints 1, 3, and 5, curve E has measurements at Timepoints 1, 2, and 3. The shorter study length for curve E requires a sample size of 950 for a power of .80, as compared with a sample size of 600 for curve C.

Treatment effect size. Figure 8 considers different effect sizes, ranging from small to medium: 0.20, 0.30, 0.40, and 0.50. In standard deviation units of the control group growth factor, the corresponding treatment-group growth rate mean increases over the control-group growth rate mean correspond to .41, .61, .82, and 1.02, respectively. The power curves again correspond to the detection of the growth rate mean difference in the five-timepoint growth model. For example, it is seen that the sample size needed for a power of 0.80 decreases from 525 to 130 as the effect size increases from 0.20 to 0.40.

Figure 9 and Figure 10 give curves corresponding to those of Figure 8 but when using only three and four timepoints, respectively. As before, the effect sizes refer to Timepoint 5.

Treatment Interactions

The final set of power curves are given for the case where there is an interaction between the treatment and the level of the initial status factor in their influence on the added growth rate factor in the treatment population. The same linear growth model as above is considered except that for the treatment population, the added growth rate factor is regressed on the initial status factor. The intercept and residual variance parameters of this regression are chosen so that the mean and variance of the added growth rate factor is the same as above. In line with the no-interaction cases above, the effect size for the interaction is considered in terms of the manifest variables at Timepoint 5 (see also Aiken & West, 1991). The slope is chosen so that an initial status factor value of one standard deviation away from its zero mean results in a certain Timepoint 5 effect size for y. These effect sizes will also be expressed in latent variable terms of how much an initial status factor value of one standard deviation away from its zero mean changes the conditional mean of the added growth factor in standard deviation units of this added growth factor. The parameter values, mean vectors, and covariance matrices for this case are given in a technical appendix, which may be obtained from Bengt O. Muthén and Patrick J. Curran.

In terms of the conceptualization of the treatment–initial status interaction, power can refer to several aspects of the treatment effect. As discussed earlier, we may consider effects expressed by one or more of the three parameters in the regression of the added growth rate factor on the initial status factor in the treatment population. The intercept represents the overall (main) treatment–control difference, the slope represents the interaction of treatment and initial status, and the residual variance represents the treatment variance increment in the growth rate that is unrelated to the interaction. Here, we limit attention to the power of detecting a nonzero slope representing the interaction. This means that we are assessing the power of detecting an interaction while allowing for a possible overall (main) treatment effect.

Figure 11 shows the power curves for the interaction effect for the five-timepoint linear growth model with manifest-variable effect sizes 0.20, 0.25, 0.30, 0.35, and 0.40 (an effect size larger than .40 was not possible given the previous choice of variance for the added growth factor). Using the above definition of latent-variable effect size, this corresponds to effect sizes .14, .18, .22, .26, and .30. While a manifest-variable interaction effect size of .40 requires a total sample size of about 275 (see curve A), a small manifest-variable interaction effect size of .20 requires over 1,000 observations (see curve E). It is seen that considerably larger sample sizes are needed to obtain a power of 0.80 than in the corresponding no-interaction models of Figure 8. This finding is in line with the case of interactions in multiple regression (Aiken & West, 1991).

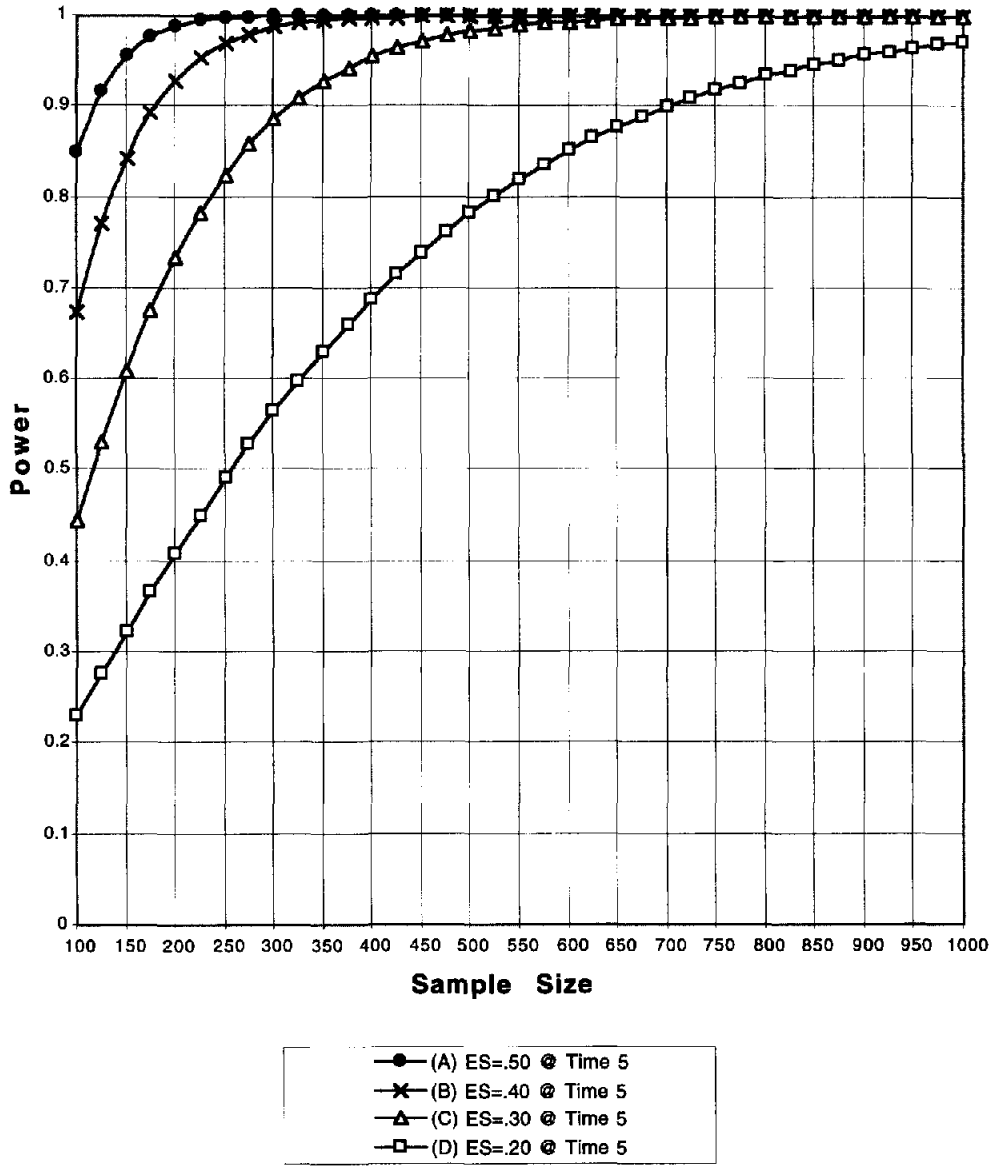

Figure 8. Power to detect various effect sizes assessed at Time 5 based on the first five measurement occasions.

Balanced Versus Unbalanced Data

We finally consider effects on power of deviations from balanced data. Here, we return to the no-interaction case considered earlier. Figure 12 shows how the power varies as a function of the proportion of treatment-group observations for a given total sample size of 250, 500, 750, and 1,000. The five-timepoint model with no interaction effect and a small effect size is considered. It is seen that the power curves are not completely symmetric around the balanced case where the proportion is .5. Choosing an unbalanced design in favor of more treatment observations is better than choosing an unbalanced design in favor of more control observations. This is because the present growth model has larger variances in the treatment group than in the control group, whereas the reverse would hold if the treatment group variances were smaller. The reverse situation was verified by using a growth model with a negative interaction slope that induced lower treatment group variances after treatment. Figure 12 also shows that the impor-

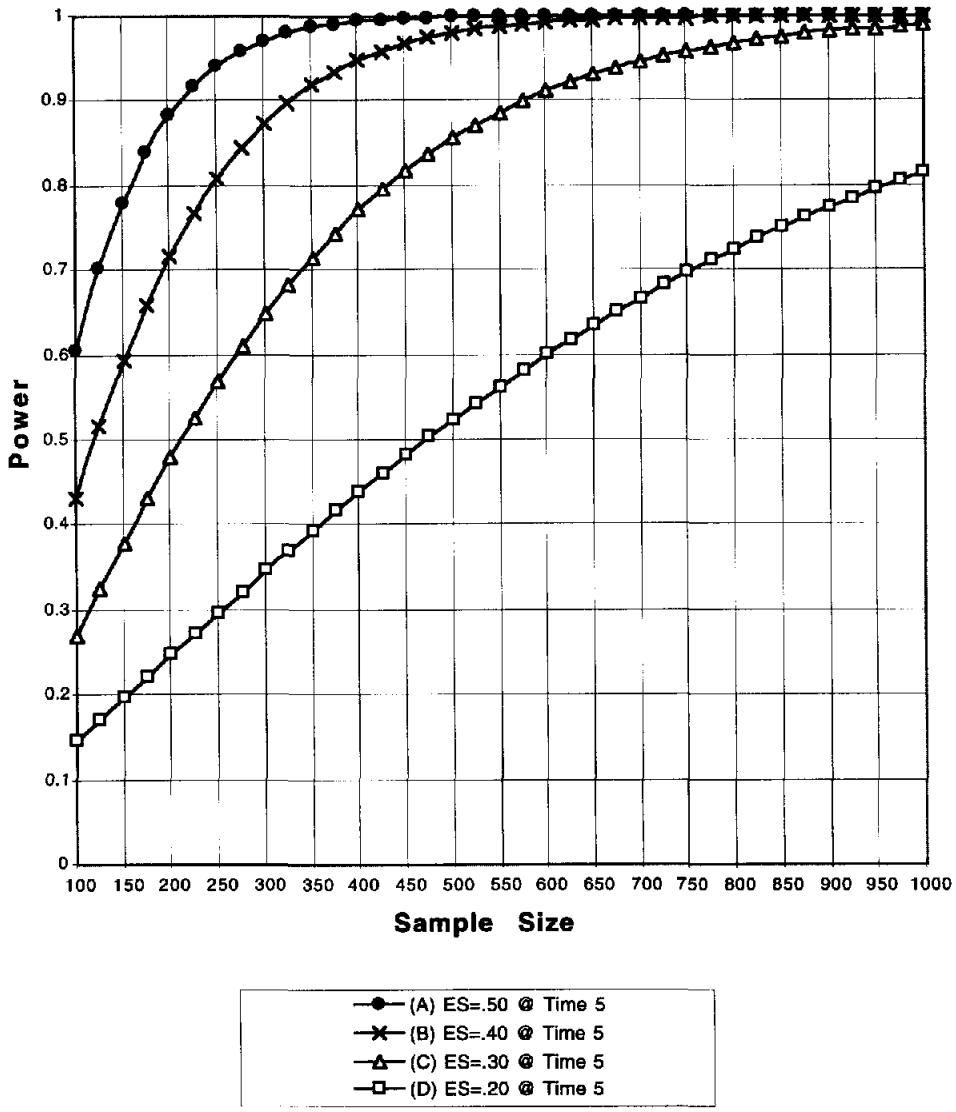

Figure 9. Power to detect various effect sizes assessed at Time 5 based on the first three measurement occasions.

tance of a nearly balanced design is more critical when the sample size, and therefore the power, is lower.

Figure 13 shows how the power varies as a function of the number of control group observations for a given number of treatment group observations and vice versa. Again, the no-interaction, five-timepoint model with a small effect size is considered. Holding the treatment group sample size fixed at 250, it is seen that the power increases rapidly as the control group sample size approaches the treatment group sample size, but that further increases in the control group sample size give quickly diminishing returns. Oversampling the treatment group has a better payoff in terms of power.

Power curves related to the proportion of treatment and control group cases are useful in terms of cost considerations. Treatment group observations are presumably considerably more expensive than control group observations. The power calculations can for example be used to answer the question: If we attempt to reduce cost by reducing the treatment group sample size from 250 to 200, how much does the control group sample size have to increase to maintain the same power? The answer from the above model is that the control group sample size has to increase by 200

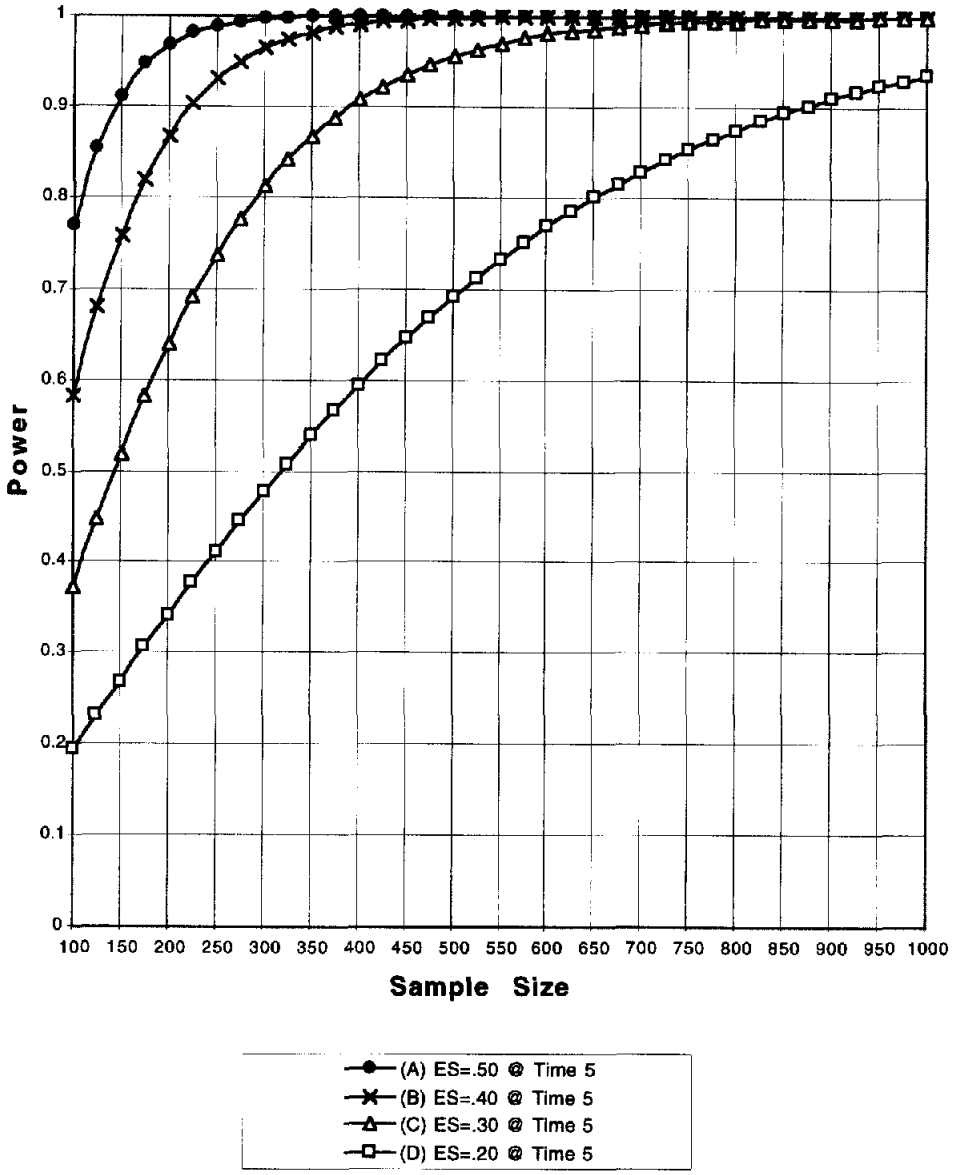

Figure 10. Power to detect various effect sizes assessed at Time 5 based on the first four measurement occasions.

from 250 to 450, implying that treatment observations are in this case four times more valuable from the point of view of power.

Real Data: Analysis and Power Estimation

A school-based preventive intervention study will be used to illustrate the general growth modeling and power estimation capabilities of the latent variable framework. The data are from a longitudinal study of Baltimore public school children in Grades 1–6 (see, e.g., Kellam, Rebok, Ialongo, & Mayer, 1994). The outcome variable that we consider corresponds to teacher-reported behavioral assessments of aggressiveness for each child in his or her class. Teacher ratings of aggression were made using the Teacher Observation of Classroom Behavior—Revised (TOCA–R) instrument. The TOCA–R measures the frequency of 18 types of aggressive behavior, each measured on a six-point scale ranging from *almost never* to *almost always.* The intervention involved a classroom team-based behavior management strategy promoting good behavior, the good behavior

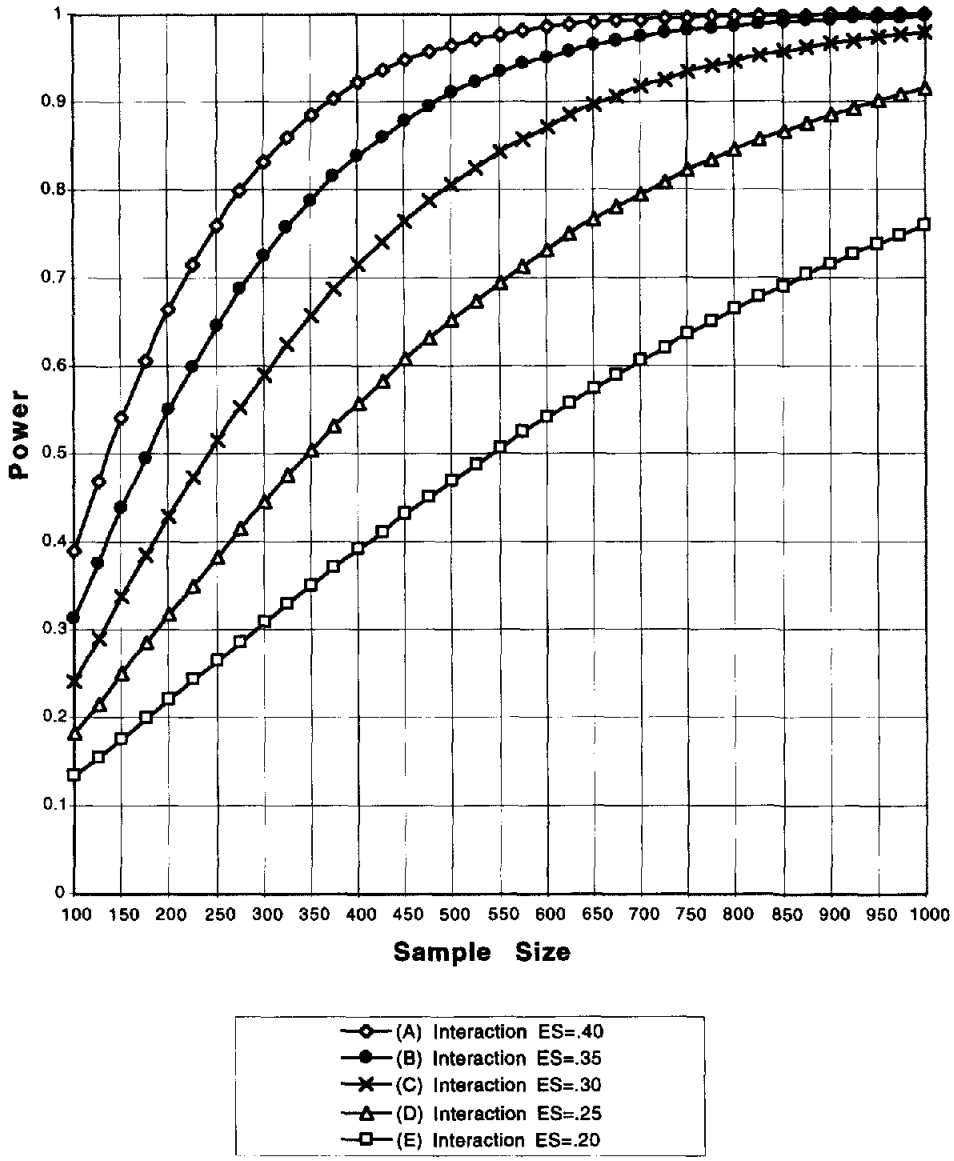

Figure 11. Power to detect an interaction given a main effect of $ES = .20$ assessed at Time 5.

game (GBG). After an initial assessment in fall of first grade, the interventions were administered during the first two grades. Assessments were made fall and spring for the first two grades and every spring thereafter through Grade 6.

Kellam et al. (1994) concluded that boys who were found to be more aggressive at the initial measurement occasion in the fall of Grade 1 benefitted more from the GBG treatment in terms of the Grade 6 outcome. This finding was obtained by ANCOVA and subsetting of the sample with respect to the initial level of aggression. We reanalyze these data using the latent curve model and allowing for an interaction between the treatment and the initial status factor. The maximum-likelihood estimator of Equation 7 will be used. This analysis differs in two important respects from the Kellam et al. analyses: using all eight timepoints instead of only the first and last, and using the latent curve model's initial status factor as covariate instead of the first timepoint measure.

We will use data from the 186 boys who were in the same intervention condition for two years. The 75 children of the GBG group are viewed as our treatment group, whereas the remaining 111 children are

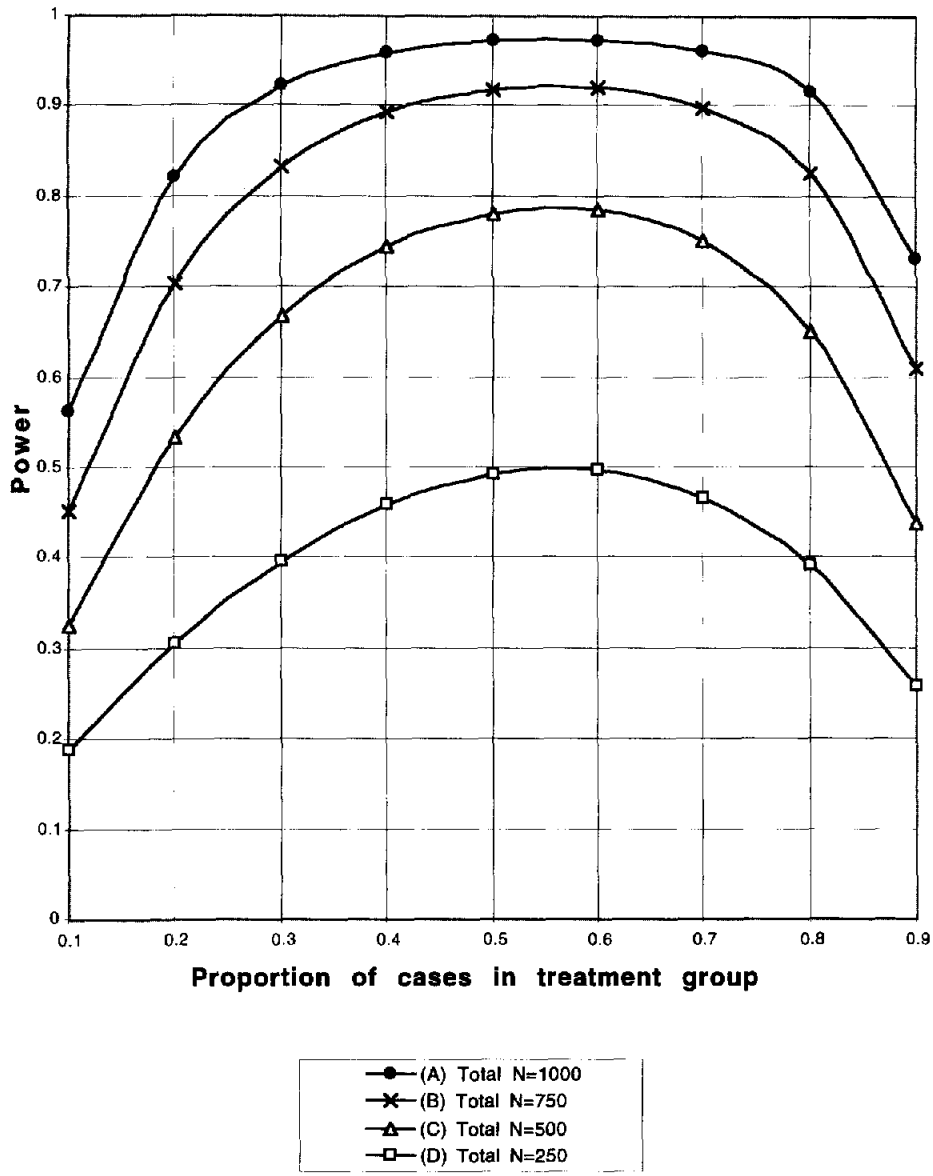

Figure 12. Power to detect a main effect of $ES = .20$ assessed at Time 5 as a function of the proportion of cases in treatment group.

viewed as our control group. The data are therefore unbalanced. Although the analyses have realistic features, the inferences given below should only be viewed as illustrative for two reasons. First, the sampling design results in clustering of individual observations within classrooms and schools and for simplicity this has been ignored in our analyses. Muthén and Satorra (1995) discussed effects of clustering when this is ignored in latent variable modeling and suggested methodology that can incorporate such sampling features. Their results show that with large average cluster size or large intraclass correlations, ignoring the clustering may lead to distorted results in terms of underestimated standard errors and likelihood-ratio chi-square values that are inflated. In the present data, the major clustering effect is most likely because of classrooms; there are eight treatment classrooms and 13 control classrooms. Given that the average class size in the analyses is only around 9 and that the aggression intraclass correlations for classrooms are not likely to be large, the Muthén and Satorra results suggest that little distortion is to be ex-

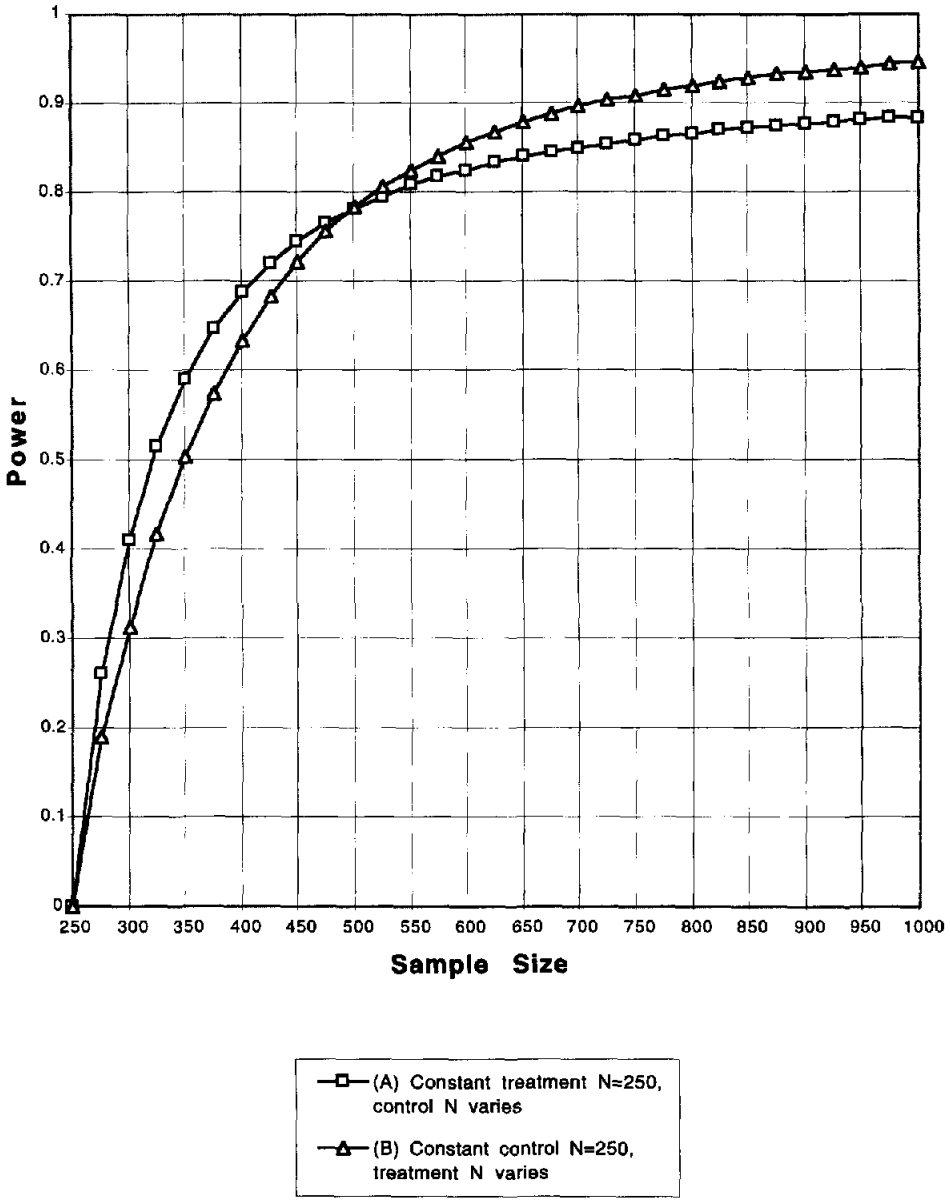

Figure 13. Power to detect a main effect of $ES = .20$ assessed at Time 5 as a function of treatment and control group sample size.

pected. Second, as is typical in longitudinal studies, there are missing data for many individuals at several of the eight timepoints. For simplicity in the present growth analyses, the missing data issue will be avoided here by using imputed values. The resulting data do, however, show the same essential features and analysis results as those of Kellam et al. (1994).

To serve as a comparison with the growth model analyses, ANCOVA was also carried out using the last timepoint as outcome and the first timepoint as the baseline covariate. This analysis indicated a significant interaction between treatment and baseline—the t value for the interaction was −2.03, i.e., $F(1, 182) = 4.12, p = 0.044$.

Latent Curve Analysis

Plots of the means of aggression over the eight timepoints are shown in Figure 14A and B for the control (CON) group and the treatment (GBG) group,

overall and based on low, medium, and high values at baseline (Timepoint 1). For the control group, there is an approximate linear progression over time overall, and there are no apparent differences in this trend among the three groups on the basis of baseline values. For the treatment group, a linear trend is also indicated, except for the boys with high baseline values, where a downturn is seen after spring of third grade. The overall linear trend in the treatment group appears similar to that of the control group. Plots of individual values show similar trends. In the latent curve analyses that follow, model fit will be evaluated using both the likelihood-ratio chi-square and the root mean squared error of approximation (RMSEA; Steiger & Lind, 1980) augmented with a 95% confidence interval calculated using the computer program

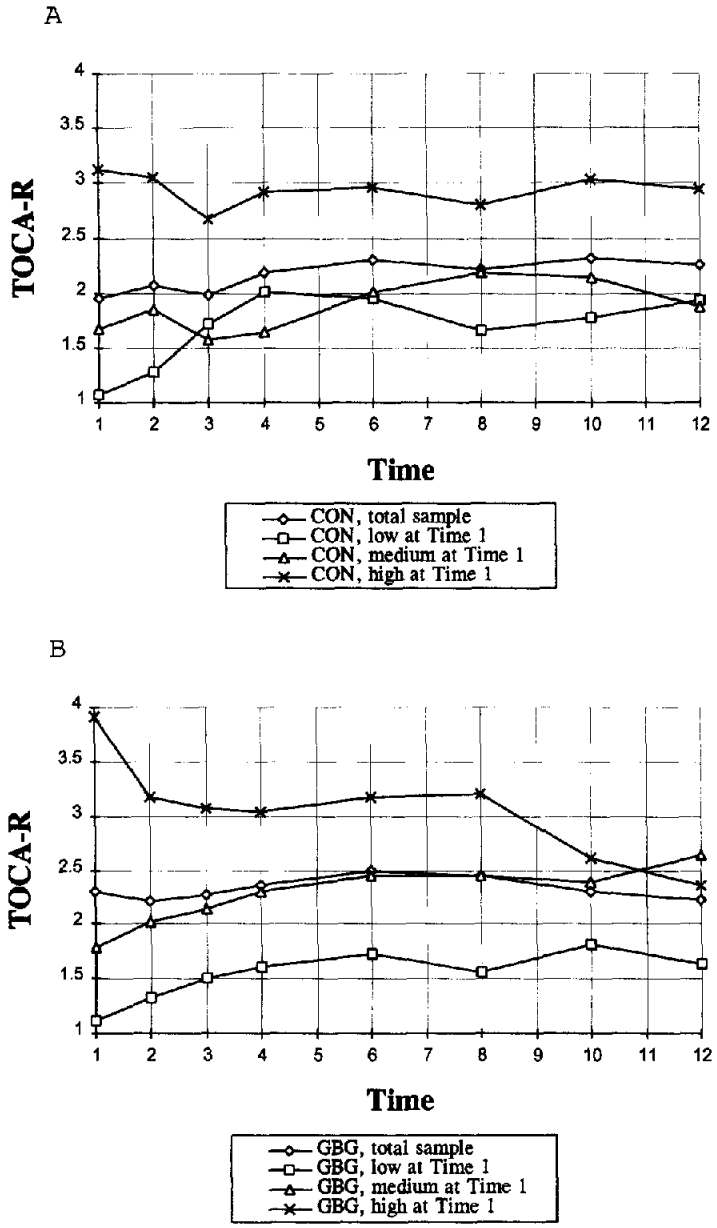

Figure 14. Panel A: Growth of Teacher Observation of Classroom Behavior—Revised (TOCA–R) scores in control group as a function of Time 1 standing. Panel B: Growth of TOCA–R scores in treatment group as a function of Time 1 standing. CON = control group; GBG = good behavior game group.

FITMOD (Browne, 1991). The RMSEA is bounded between zero and infinity and values falling below about 0.05 are thought to reflect "close" model fit (Browne & Cudeck, 1993).

Step 1: Control group analysis. In line with our proposed analysis strategy, the control group was analyzed in a first step. A two-factor, linear growth model was chosen using the x_t scores 0, 1, 2, 3, 5, 7, 9, and 11 to capture the fact that only spring measures were collected after the first two grades. The residuals were allowed to have unequal variances and be correlated at adjacent timepoints, $\chi^2(24, N = 111) = 31.9, p = 0.13$, RMSEA = .055, confidence interval (CI) (0, 11). Here, the positive growth rate factor mean and variance are both significantly different from zero at the .01 level. Evidence of nonlinear growth was also investigated. A quadratic factor was added using squared x_t scores 0, 1, . . ., 121. It should be recognized that the use of a quadratic function is only relevant for a limited time period of the development. In these analyses, the x_t scores were not centered at their means as is sometimes advocated when using a quadratic function. This is for two reasons. First, centering changes the interpretation of the initial status factor to status at the centering point so that in the treatment group, it is no longer defined as a pre-intervention status factor. Because of this, the two-group model approach no longer has the desired feature of describing the influence of pre-intervention status on the change in development. Second, the customary problem of near collinearity of x_t scores when not centering has not been found to be an issue in these latent curve analyses. In the latent variable framework, these scores do not appear as predictor variables but as coefficients (parameters) and there is an issue of high parameter estimate correlations. In single-group analysis for the control group and for the treatment group, the uncentered and centered model versions have exactly the same model fit and only involve a reparameterization. Comparison of the uncentered and centered versions of the single-group model for the control group and for the treatment group did show a high correlation between the parameter estimates of the means of the linear and quadratic factors for the uncentered solutions (−.93 and −.95, respectively) but not for the centered solutions. This high correlation is, however, expected and does not involve parameters for which a clear-cut separation is key. Furthermore, the high correlation does not appear to give rise to any problems of numerical instability such as inflated standard errors. For example, the parameterizations for the uncentered and centered solutions can be shown to give the same quadratic factor mean and for this parameter the estimates and the estimated standard errors are exactly the same for both the control and treatment group analysis when comparing the uncentered and centered solutions. The analysis results (uncentered model) were as follows. The quadratic factor showed no significant variation across individuals but did obtain a negative mean which was significant at the 1% level; with the quadratic variance fixed at zero, $\chi^2(23, N = 111) = 24.97, p = 0.35$, RMSEA = .028, CI(0, .093). In summary, the final, nonlinear control group model indicates a linear increase in the aggression score over grades with a slight decrease at the later grades.

Step 2: Treatment group analysis. As a second step, the treatment group was analyzed separately. The quadratic model found in the control group obtained a good fit also here; $\chi^2(23, N = 75) = 30.66, p = 0.13$, RMSEA = .067, CI(0, .13). Attempts to allow for a nonzero variance in the quadratic factor failed because of nonconvergence. The analysis shows a quadratic factor mean that is not significantly different from zero at the 5% level (the t value is −1.60). This implies that a linear growth model fits well in the treatment group. The estimated mean for the quadratic factor, while insignificant, is the same as in the control group. Given the low treatment group sample size of 75, the power to detect a nonzero quadratic mean at the estimated parameter values is only .34 as estimated by the Satorra–Saris (1985) method (a treatment group sample size of 220 would be required for a power of .80).

Step 3: Two-group analysis without interactions. As a third step, the control and treatment groups were analyzed simultaneously in a two-group analysis. Here, the above quadratic growth model was used for the control group and the parameters for these three factors were held equal across the two groups. Although there is not statistically significant evidence of a quadratic factor in the growth analysis of the treatment group, the two-group approach includes this factor. Note, however, that this factor is specified to have zero variance (and zero covariances) and only contributes a mean parameter. In the absence of a priori theory, the added effects of treatment are modeled based on the impressions of the plots above. In the treatment group, the added growth factor is chosen to be linear for simplicity. Preliminary analyses indicated that there was no variation across individuals for this factor and the corresponding variance is therefore

fixed at zero. As a first analysis, no interaction is allowed for. This two-group model fit the data reasonably well, $\chi^2(51, N = 186) = 75.00$, $p = .02$, RMSEA = .050, CI(.013, .078). The treatment effect is here described by the mean of the added linear factor. The estimate of this mean is, however, not significant even at the 10% level.

Step 4: Two-group analysis with interactions. As a second two-group analysis, the initial status factor in the treatment group is allowed to influence the added linear factor, thereby accommodating a possible interaction between treatment and initial status. The added linear factor was found to have zero residual variance so that its variation is solely determined by the initial status variation. Adding a single slope parameter to represent the interaction, this two-group analysis resulted in a well-fitting model, $\chi^2(50, N = 186) = 64.56$, $p = 0.08$, RMSEA = .04, CI(0, .07). The interaction is significant (the one degree of freedom chi-square difference value is 10.44 with $p = .002$). It is noteworthy that no treatment effect would have been discovered if the interaction effect had not been included.

Step 5: Sensitivity analysis of final model. To test for deviations from successful randomization into treatment and control groups, the above two-group (50 degree of freedom) model was relaxed to allow for the initial status factor mean of the treatment group to deviate from the zero value of the initial status factor in the control group. This test indicated that the treatment group mean was marginally higher than zero (the chi-square difference with one degree of freedom was 4.41 with $p = 0.04$). Even when allowing for this pre-existing difference, however, the interaction effect remained significant at the same level and the more parsimonious model of no pre-existing differences was maintained.

Finally, the overall effect of treatment on the growth factors was tested. Here, the two treatment effect parameters were set at zero, resulting in a significant worsening of the fit relative to the 50 degree of freedom model (the chi-square difference test value with 2 *df* was 12.01 with $p < 0.01$). The hypothesis of no treatment effect is therefore rejected.

Estimated two-group model. Table 1 presents the estimates of the two-group treatment interaction model with 50 degrees of freedom. The model specification is given in a technical appendix, which may be obtained from Bengt O. Muthén or Patrick J. Curran. The variances for the initial status and linear growth rate factors are significantly different from zero, indicating across-student heterogeneity in the across-grade trajectories of aggressive behavior. There is a significant treatment effect on growth that interacts with initial status. The interaction is expressed in terms of the regression of the added linear growth factor on the initial status factor, which has a significant negative estimate of the slope. The 95% confidence interval for the slope is −.081, −.023. The negative slope indicates an interaction of the expected kind: Initially more aggressive boys benefit more from the intervention. Given that the residual variance is zero in this regression, the variation in the added linear growth factor of the treatment group is solely due to the variation in initial status. The negative intercept is not significant, indicating that we cannot reject that there is no overall treatment effect. The 95% confidence interval for the intercept is −.038, .006. It is interesting to compare the *t* value for the two-group growth model interaction slope with that of the ANCOVA interaction. While the former is −3.35 with $p = .003$, the latter is only −2.03 with $p = .044$.

Figure 15 displays the estimated model in terms of the model implied means for the outcome variables. Here, graphs are given for the control and treatment groups at three different levels of the initial status factor: at the mean and at half a standard deviation below and above the mean. The figure shows the effect of the interaction so that only individuals at higher initial status values benefit clearly from the treatment. At the last timepoint, the mean difference between the control and treatment group individuals in the highest initial status category of the figure is about one third of a standard deviation.

It is interesting to note that the treatment effect findings were made possible by using a two-group approach with an added growth factor to capture treatment effects. A quadratic growth factor could not be found in the separate analysis of the treatment group. In contrast, the two-group analysis revealed an interaction such that a nonlinear trajectory with a stronger downward trend at later grades is realized for the more aggressive subset of boys. Furthermore, a conventional two-group growth analysis does not give an equally clearcut analysis. A conventional two-group analysis would use the same three growth factors in both groups and study differences between the parameter estimates of the linear and quadratic growth factors. A covariance between the initial status factor and the quadratic factor would not be included if the latter does not have a significant variance, which is the case here. This means that the same quadratic model as

Table 1
Final Two-Group Growth Model for Aggression Data

Parameter	Control group (n = 111)	Treatment group (n = 75)	Both
Growth factors			
Initial status			
M			0.[a]
Variance			0.80 (0.1100)
Linear growth rate			
M			0.0860 (0.0200)
Variance			0.0045 (0.0012)
Quadratic growth rate			
M			−0.0051 (0.0016)
Residual variance			0.[a]
Added linear growth rate regressed on initial status			
Intercept		−0.016 (0.013)	
Slope		−0.052 (0.015)	
Residual variance		0.[a]	
Growth factor covariances			
Initial-status linear growth rate			−0.0015 (0.0089)
Residual variances for outcome variables			
Time 1	0.44 (.088)	0.53 (.14)	
Time 2	0.45 (.079)	0.44 (.12)	
Time 3	0.41 (.069)	0.50 (.11)	
Time 4	0.52 (.080)	0.70 (.13)	
Time 5	0.51 (.079)	0.74 (.13)	
Time 6	0.42 (.074)	0.80 (.15)	
Time 7	0.26 (.083)	0.24 (.10)	
Time 8	0.29 (.094)	0.61 (.18)	
Residual covariances for outcome variables			
Time 1–2	0.29 (.071)	0.070 (.100)	
Time 2–3	0.077 (.031)	0.031 (.061)	
Time 3–4	0.25 (.058)	0.032 (.097)	
Time 4–5	0.13 (.047)	0.099 (.069)	
Time 5–6	−0.0059 (.049)	0.031 (.11)	
Time 6–7	0.055 (.058)	0.021 (.072)	
Time 7–8	−0.11 (.066)	−0.032 (.11)	
Common intercept for the outcome variables			2.04 (0.023)

Note. Standard errors are given in parentheses; χ^2 (50, N = 186) = 64.56, p = 0.08.
[a] Parameter is fixed in this model.

was found above for the control group is applied in both groups. This two-group model fits the data well, $\chi^2(47, N = 186) = 59.44, p = .11$, RMSEA = .038, CI(0, .069). It does not, however, show significant group differences in terms of t values for either the linear or quadratic means and the linear mean estimate is, in fact, higher for the treatment group. In conclusion, these illustrative analyses indicate that an intervention effect is only seen for initially more aggressive students so that an interaction effect is present with no overall, main effect.

Power Estimation

We finally consider power estimation for the final two-group model of Table 1. On the basis of this estimated model, the effect size for the interaction is .38 using the same definition as given earlier for the artificial data. Figure 16 gives the power of detecting a nonzero interaction slope parameter in a model with parameter values equal to the estimated values in Table 1. The three curves in this figure indicate how small the combined sample size could have been

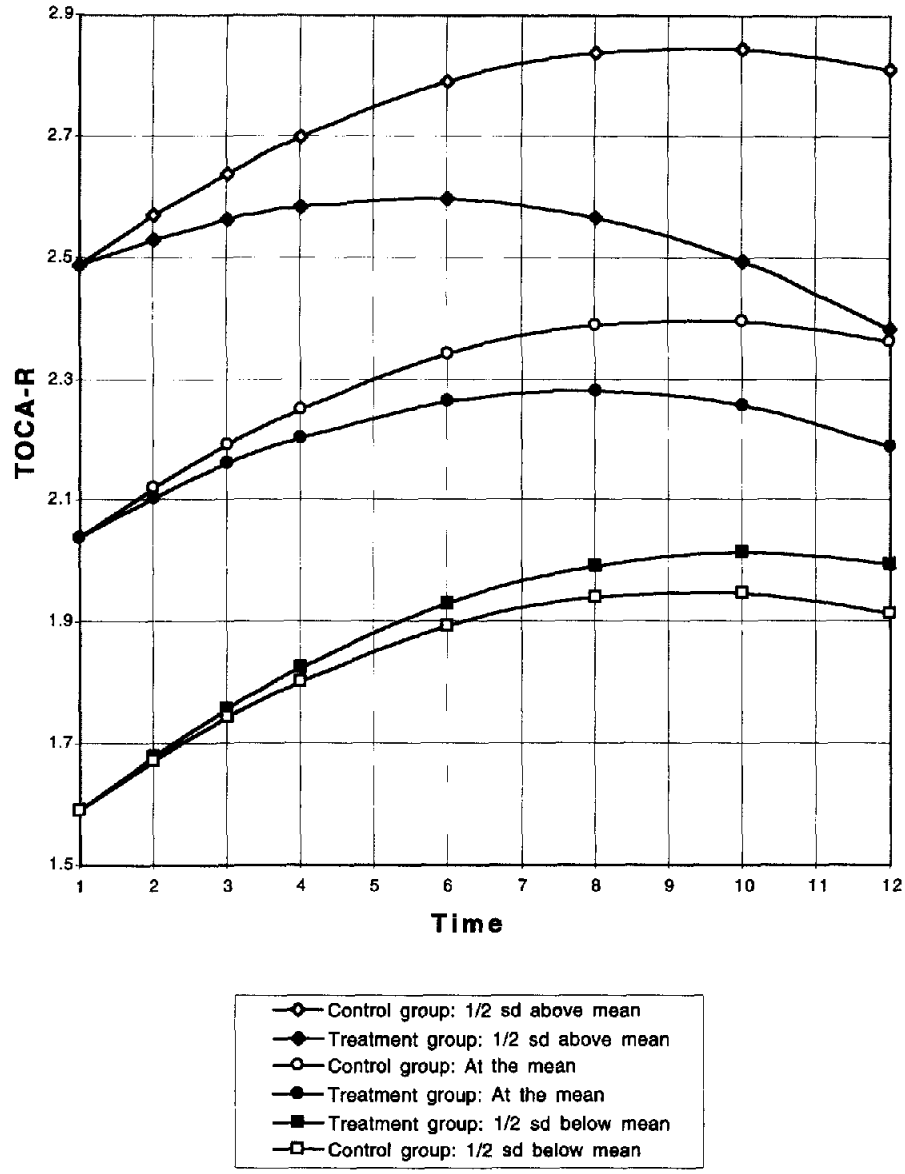

Figure 15. Model implied growth trajectories of Teacher Observation of Classroom Behavior—Revised (TOCA-R) scores as a function of initial status. Each timepoint represents one 6-month interval.

while still making it possible to find the treatment interaction effect. Curve A gives the power as a function of total sample size with balanced data. Curves B and C show the power with the treatment group sample size fixed at its actual value while varying the control group sample size and with the control group sample size fixed at its actual value while varying the treatment group sample size, respectively.

With the actual sample of 75 and 111 in the treatment and control group, respectively, the power to detect the interaction is .90. This shows that an interaction can be detected with high power even for a total sample size of only 186 when the interaction effect size is moderate (.38 in this case). For balanced data, curve A shows that a total sample size of 130 instead of the original 186 observations would have been sufficient to detect the interaction with a power of .80. If instead a balanced design had been used with the original total sample size of 186, the power would have been .92 instead of .90.

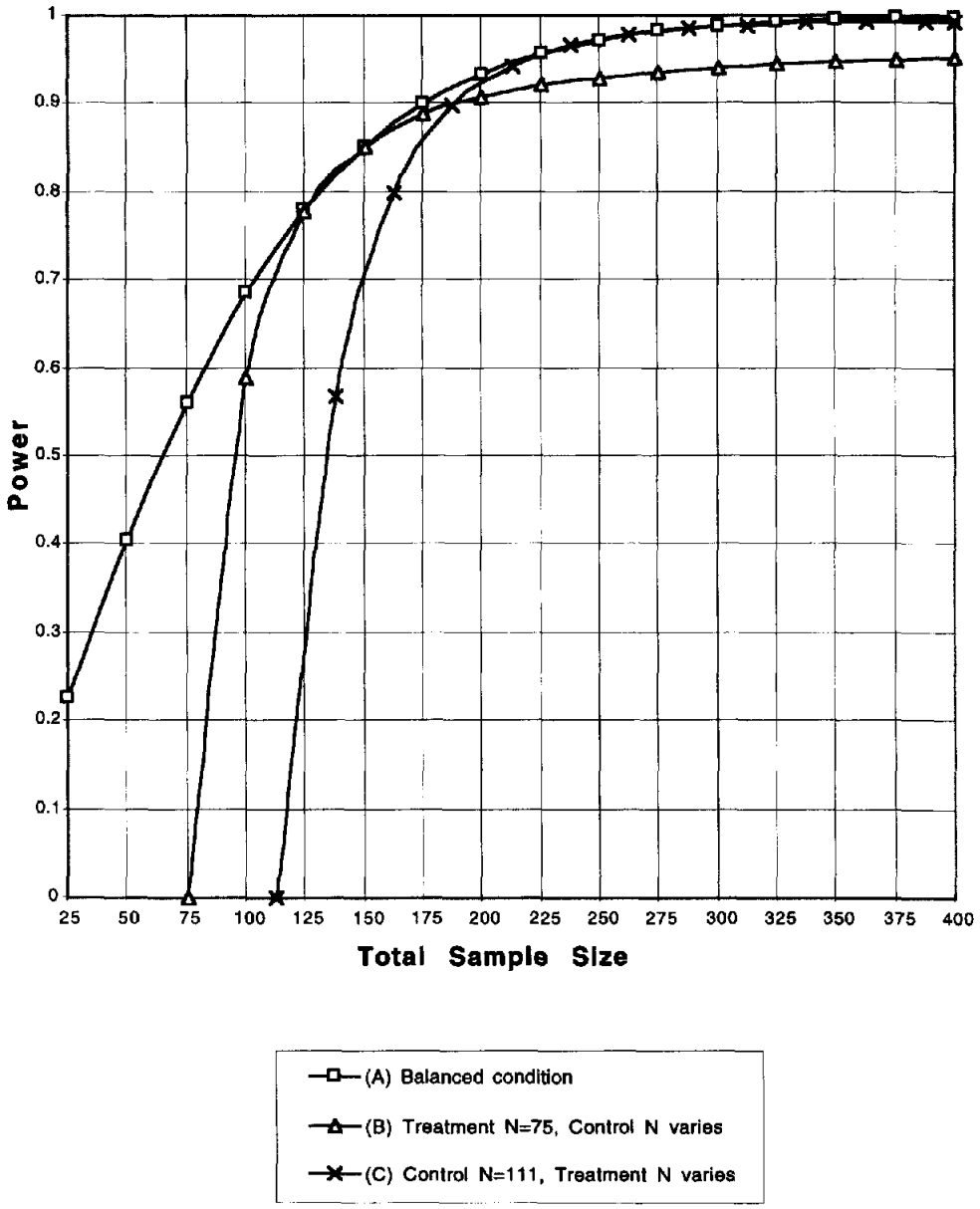

Figure 16. Power to detect the interaction treatment effect from the Kellam et al. (1994) aggressive behavior intervention.

Conclusions

The analyses of this paper indicate some of the generality of the latent variable approach to longitudinal modeling of individual differences in development. In the specific application of multiple-population analysis of intervention effects, it is possible to separate normative development in a control group from the change in course due to treatment. Change in course due to the interaction between treatment and initial status can be captured by structural regressions between latent curve variables. Power estimation is readily available through standard latent variable techniques. The generality of the latent variable approach to longitudinal modeling and power estimation is particularly exciting when considering that the multiple-population study of interventions can be put into the framework of either one of the longitudinal modeling generalizations shown in Figures 2–4. To this may also be added the generalization of

multiple indicators of latent variable constructs. The general modeling potential is largely unexplored and hopefully this article can stimulate new types of analyses.

The article focuses on multiple-population longitudinal studies as motivated by an intervention context. It is clear, however, that these analysis and power estimation techniques are more generally applicable. The multiple-population settings may, for example, involve gender differences or differences among different risk populations. Single-population settings also benefit from power estimation for detection of certain developmental patterns.

The artificial data studies of power show the importance of going beyond the ANCOVA approach and use more than two timepoints in assessing intervention effects. They also illustrate the tradeoffs between using more timepoints or larger samples. They show that interaction effects can be detected without unduly large sample sizes if the interaction effects are sizable. Furthermore, the power calculations show that designs that have a balance between control and treatment group sizes are not always the most powerful.

It is clear that power estimation is directly related to the parameter values of a specific model and therefore the above power curves are only illustrative. The importance of the approach is, however, that the researcher can compute his or her own power curves for a model with parameter values that he or she hypothesizes. Many different scenarios can be easily assessed and can give important guidance for design decisions.

The real-data analyses of aggressive behavior among elementary school children showed the complexities of a real intervention analysis. The intervention effect here is only seen for initially more aggressive students so that an interaction effect is present with no overall, main effect.

Although complex, the approaches discussed above rely on strongly simplified assumptions. A central issue of growth modeling that has been ignored here is missing data, in particular attrition over time. Attrition should also play a key role in design decisions. While the use of many timepoints increases power, this benefit is reduced by an increasing attrition rate. Also, the above discussion focuses on normally distributed data for the outcome variables, while many intervention studies have strongly nonnormal and categorical outcome variables. Recent research (Curran, West, & Finch, 1996; Satorra & Neudecker, 1995) shows that there is a strong reduction in power when variables clearly deviate from normality. Furthermore, longitudinal data are often obtained through cluster sampling giving rise to multilevel data. This fact has also been ignored above, and analysis methods and power calculation need to be studied for such situations.

References

Aiken, L. S., & West, G. W. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Arminger, G., & Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association, 85,* 195–203.

Bentler, P. M., & Chou, C. P. (1988). Practical issues in structural equation modeling. In J. S. Long (Ed.), *Common problems/proper solutions: Avoiding error in quantitative research.* Newbury Park, CA: Sage.

Bock, R. D. (1989). *Multilevel analysis of educational data.* San Diego, CA: Academic Press.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Brown, H., Kellam, S., & Liao, J. (1996). *Principles for designing randomized preventive trials in mental health.* Manuscript submitted for publication.

Browne, M. W. (1991). *FITMOD: A computer program for calculating point and interval estimates of fit measures.* Unpublished manuscript.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & K. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Browne, M. W., & DuToit, S. H. C. (1991). Models for learning data. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change. Recent advances, unanswered questions, future directions* (pp. 47–68). Washington DC: American Psychological Association.

Bryk, T., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Bryk, A. S., & Weisberg, H. I. (1976). Value-added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics, 1,* 127–155.

Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education, 8,* 158–233.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New Jersey: Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design & analysis issues for field settings.* Boston: Houghton Mifflin.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulations of questions, design and analysis* (Tech Rep). Stanford, CA: Stanford Evaluation Consortium.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington.

Curran, P. J. (1994). *The robustness of confirmatory factor analysis to model misspecification and violations of normality.* Unpublished doctoral dissertation, Arizona State University.

Curran, P. J., Harford, T. C., & Muthén, B. (1996). The relation between heavy alcohol use and bar patronage: A latent growth model. *Journal of Studies on Alcohol, 57,* 410–418.

Curran, P. J., Stice, E., & Chassin, L. (1997). The relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology, 65,* 130–140.

Curran, P. J., West, S. G., & Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

Diggle, P. J., Liang, K. L., & Zeger, S. L. (1994). *Analysis of longitudinal data.* Oxford, England: Clarendon Press.

Goldstein, H. I. (1987). *Multilevel models in educational and social research.* London: Oxford University Press.

Goldstein, H. I. (1995). *Multilevel statistical models.* London: Edward Arnold.

Jaccard, J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression: Vol. 114. Quantitative Applications in the Social Sciences Series.* Thousand Oakes, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt Books.

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry, 35,* 963–974.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38,* 963–974.

Longford, N. T. (1993). *Random coefficient models.* Oxford, England: Clarendon Press.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

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

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

Meredith, W., & Tisak, J. (1984, June). *Tuckerizing curves.* Paper presented at the Psychometric Society annual meeting, Santa Barbara, CA.

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

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics, 22,* 48–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132.

Muthén, B. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model. Theoretical integration and user's guide.* Mooresville, IN: Scientific Software.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations: Presidential address to the Psychometric Society, July, 1989. *Psychometrika, 54,* 557–585.

Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change. Recent advances, unanswered questions, future directions* (pp. 1–17). Washington, DC: American Psychological Association.

Muthén, B. (1993). Latent variable modeling of growth with missing data and multilevel data. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp. 199–210). Amsterdam: North Holland.

Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22,* 376–398.

Muthén, B. (1995). *Longitudinal studies of achievement growth using latent variable modeling.* Unpublished technical report, University of California, Los Angeles.

Muthén, B. (1996). Growth modeling with binary responses. In A. V. Eye & C. Clogg (Eds.), *Categorical variables in developmental research: Methods of analysis* (pp. 37–54). San Diego, CA: Academic Press.

Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel data. In A. E. Raftery (Ed.), *Sociological methodology* (pp. 453–480). Washington, DC: Blackwell.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural

equation modeling with data that are not missing at random. *Psychometrika, 42,* 431–462.

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 267–316). Washington, DC: American Sociological Association.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics, 14,* 1–17.

Raundenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U.S. high-school data. *Journal of Educational Statistics, 16,* 295–330.

Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171–209). New York: Springer.

Rogosa, D. R. (1991). A longitudinal approach to ATI research: Models for individual growth and models for individual differences in response to intervention. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 221–248). Hillsdale, NJ: Erlbaum.

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

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models.* Newbury Park, CA: Sage.

Saris, W. E., & Stronkhorst, L. H. (1984). *Causal modeling in nonexperimental research: An introduction to the LISREL approach.* Amsterdam: Sociometric Research Foundation.

Satorra, A., & Neudecker, H. (1995, July). *The asymptotic distribution of the chi-square goodness of fit test statistic under distributional and model misspecification.* Paper presented at the annual meeting of the Psychometric Society, Leiden, The Netherlands.

Satorra, A., & Saris, W. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 51,* 83–90.

Seltzer, M., Frank, K., & Bryk, A. (1994). The metric matters: The sensitivity of conclusions concerning growth in student achievement to choice of metric. *Educational Evaluation and Policy Analysis, 16,* 41–49.

Sörbom, D. (1982). Structural equation models with structured means. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 183–195). Amsterdam: North Holland.

Steiger, J. H., & Lind, J. M. (1980, June). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research, 29,* 263–288.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23,* 19–23.

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

Received February 14, 1996
Revision received October 30, 1996
Accepted February 10, 1997 ■