

A Structural Equation Model for Measuring Residualized Change and Discerning Patterns of Growth or Decline

Tenko Raykov

University of Melbourne

This paper is concerned with two theoretically and empirically important issues in longitudinal research: (1) identifying correlates and predictors of change and (2) discerning patterns of change. Two traditional methods of change measurement—the residualized observed difference and the residualized gain score—are discussed. A general structural equation model for measuring residualized true change and studying patterns of true growth or decline is described. This approach allows consistent and efficient estimation of the degree of interrelationship between residualized change in a repeatedly assessed psychological construct and other variables, such as studied/presumed correlates and predictors of growth or decline on the latent dimension. Substantively interesting patterns of change on the

trait level, such as regression to the mean, overcrossing, and fan-spreading, can be discerned. The model is useful in research situations in which it is of theoretical and empirical concern to identify those variables that correlate with, or can be used to predict, such patterns of true growth or decline that deviate from a group-specific trend in longitudinally-measured psychological constructs. The approach is illustrated using data from a cognitive intervention study of plasticity in fluid intelligence of aged adults (Baltes, Dittmann-Kohli, & Kliegl, 1986). *Index terms: correlates of growth/decline, fan-spreading, measurement of change, overcrossing, predictors of growth, regression to the mean, structural equations modeling, true change.*

Psychologists studying inter- and intra-individual differences often need to estimate the degree of interrelationship between change in a person's performance and additional variables. From these data, theoretically relevant and empirically interpretable correlates and predictors of growth or decline in longitudinally-measured psychological constructs can be identified. The assessment of these correlates and predictors is of special importance in psychological research for answering questions, such as "What are the characteristics of people that learn (grow) the fastest/slowest?" (e.g., Cronbach & Furby, 1970, p. 77; Rogosa, Brandt, & Zimowski, 1982). Similarly, in a geropsychological context, it is important to identify variables that are correlated with atypically accelerated or slowed decline in cognitive functioning (e.g., Baltes, Dittmann-Kohli, & Kliegl, 1986; Belsky, 1990; Cavanaugh, 1990; Woodruff-Pak, 1988).

One of the problems in the assessment of change in psychological research has been regression to the mean (e.g., Bereiter, 1963; Linn & Slinde, 1977; Lord, 1963; Nesselroade, Stigler, & Baltes, 1980). Some discussions of this phenomenon (e.g., Bereiter, 1963; Linn & Slinde, 1977; Nesselroade et al., 1980) have concentrated primarily on manifest variables. The error of measurement has been identified by these authors as a major cause of regression to the mean in the observed scores. However, changes may occur in the latent dimension under study that lead to observing regression to the mean in the observed values (e.g., Nesselroade et al., 1980). Furthermore, it is often substantively more

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 17, No. 1, March 1993, pp. 53-71

© Copyright 1993 Applied Psychological Measurement Inc.
0146-6216/93/010053-19\$2.20

insightful to study this phenomenon and other patterns of change, such as overcrossing or fan-spreading (see below), in terms of latent scores rather than observed scores.

This paper initially discusses two “classical” approaches to change measurement—the residualized observed difference score (Cronbach & Furby, 1970; Lacey & Lacey, 1962; Zimmerman & Williams, 1982a) and the residualized gain score (Tucker, Damarin, & Messick, 1966). A general structural equation model (SEQM) for measuring residualized true change and studying patterns of true change also is then described. In contrast to the classical methods, the SEQM provides, in a multiple-indicator context, consistent and efficient estimation of the degree of interrelationship between residualized true change (i.e., the true change from which true initial level is partialled out) across two measurements and other variables, such as studied/presumed correlates and predictors of growth or decline on the latent dimension of interest. (For simplicity these variables here are called “correlates.”) This approach focuses on change in persons’ true scores; therefore, patterns of change on the latent variable, such as regression to the mean, overcrossing, and fan-spreading can be tested. The structural equation method is useful in research situations in which it is of theoretical and empirical concern to identify variables that are correlated with residualized change in a repeatedly assessed psychological construct and that can be used to predict individual patterns of excessively accelerated or slowed growth or decline in the construct, as compared to a group-specific trend. The model also is useful in ascertaining the pattern of true change over time. The same logic can be used to test for regression to the mean, overcrossing, or fan-spreading with respect to observed scores. The method is illustrated using data from a cognitive intervention study of reserved capacity in fluid intelligence of aged adults (Baltes et al., 1986).

Measurement of Change, Predictors, and Correlates of Growth or Decline

Psychological data typically consist of fallible observed scores and, consequently, possibly quite unreliable difference scores in a repeated measures design (e.g., Bond, 1979; Rogosa et al., 1982; Rogosa & Willet, 1983; Zimmerman & Williams, 1982a, 1982b). Measurement error has a biasing effect on estimates of correlation coefficients, such as the correlation between change and initial level (“spurious” correlation; see e.g., Bereiter, 1963; Cronbach & Furby, 1970) or correlations between observed change and other psychological variables. The error of measurement in observed variables can also lead to inconsistent estimates in regression analysis and in the analysis of covariance (“under-adjustment” for group differences; e.g., Cochran, 1968; Isaac, 1970; Lord, 1960; Schaie & Hertzog, 1985).

Change measurement attempts to identify variables that are closely related to systematic differences in individual change (e.g., Rogosa et al., 1982). Change, however, generally is not independent of initial level (e.g., Bereiter, 1963; Bond, 1979; Jin, 1992; Rogosa & Willett, 1985; Tucker et al., 1966; Wilder, 1965). That dependence can result from the presence of measurement error or from specific changes along the true latent dimension. These changes can be caused by such factors as day-to-day variation, biological/physiological rhythm, or person reactivity. Extreme true scores can often result from a rare interaction of influential factors that is unstable over time (e.g., Jamieson & Howk, 1992; Nesselroade et al., 1980).

Several approaches have been considered in the psychometric literature for measuring change in repeatedly assessed variables (see Raykov, 1992b). The simple observed differences between a person’s scores on some measure X'_i (the first assessment; hereinafter referred to as pretest) and X''_i (the second assessment; hereinafter referred to as posttest) at two assessments is represented by $D_i = X''_i - X'_i$, $i = 1, \dots, N$, where N is the number of persons. These observed differences are unbiased estimates of true change but can have serious limitations that have been discussed extensively elsewhere

(e.g., Baltes et al., 1986; Bereiter, 1963; Burr & Nesselroade, 1990; Cronbach & Furby, 1970; Gulliksen, 1950; Linn & Slinde, 1977; Lord, 1963; Lord & Novick, 1968; Rogosa et al., 1982; Rogosa & Willet, 1983; Zimmerman & Williams, 1982b).

The residualized observed difference score— $\hat{D} = X'' - (\hat{X}'')$ —was proposed by investigators concerned about spurious correlation between change and level on the pretest (e.g., Cronbach & Furby, 1970; Lacey & Lacey, 1962; Manning & DuBois, 1962; Webster & Bereiter, 1963; in the remainder, the person subscript i will be eliminated when no ambiguity will result). Within this approach, the linear regression (\hat{X}'') of posttest on pretest is partialled out from the posttest. \hat{D} can be used to identify variables that correlate highly (positively or negatively) with a measure of change intended to account for dependence of change on pretest scores. However, because \hat{D} results from partialing out observed pretest scores from observed posttest scores this approach can lead to inconsistent estimates of individual change and, therefore, yield misleading estimates of correlations. This is of particular importance if measurement error is sizeable in the pretest, because then the assumption of an “error-free” predictor is seriously violated as the observed fallible pretest is the regressor (e.g., Cochran, 1968; Kendall & Stuart, 1962; Lord, 1960). Furthermore, this method does not take into account the error of measurement in the posttest, because the predicted posttest is subtracted from observed posttest.

In contrast to the residualized observed difference method, the base-free measurement of change partials out true pretest from true posttest (Bond, 1979; Messick, 1981; Raykov, 1992a, 1992b; Rogosa & Willett, 1985; Tucker, 1979; Tucker et al., 1966; Zimmerman & Williams, 1982a). Within the classical test theory framework (e.g., Lord & Novick, 1968), a decomposition of true change $D_T = T'' - T'$ for observed scores $X' = T' + E'$ and $X'' = T'' + E''$ (T' and T'' are true scores and E' and E'' are errors of measurement) has been postulated by Tucker et al. (1966) in the form

$$D_T = aT' + g, \quad (1)$$

where g is the residualized gain score that has been assumed to be independent of T' , a is the regression slope of true change on true pretest, and g can be correlated with additional variables to identify correlates and predictors of excessively accelerated or slowed growth or decline.

A potentially serious drawback of this approach, however, is that it requires the substitution of correlation and reliability coefficients into the correlation estimator of interest. Thus, because the correlations and reliabilities are estimated from the data, the method leads to the accumulation of estimation imprecision. Another consequence is the implied dependence of the resulting correlation estimates on the particular sample used to derive the reliability estimates. This dependence reduces the generalizability of the results (e.g., McCallum, 1986; McCallum, Roznowski, & Necowitz, 1992). Further, in the original base-free measurement of change approach, the means and their development over time were not considered.

The methods of change measurement discussed thus far were initially formulated in terms of a single repeatedly assessed variable. In psychological research, however, it is often recommended that latent constructs be measured with more than one indicator. Single measures of a latent dimension of interest generally lack reliability, because each measure addresses only one specific aspect of the underlying latent dimension it purports to measure. Therefore, each indicator taken alone contains limited, possibly even misleading, information about the unobservable construct. However, employing several interrelated indicators often provides multifaceted, converging information regarding the latent dimension.

Therefore, a major shortcoming of the classical approaches to change measurement is that they are based on a limited amount of information about change and its correlates. Not less importantly,

desirable statistical properties, such as consistency or efficiency, have not been established for either the correlation estimator obtained with the original Tucker et al. (1966) approach or for the residualized observed difference method. Also, procedures allowing for testing the underlying model assumptions have not been developed for either of these two approaches.

A General Structural Equation Model for Measuring Residualized True Change and Studying Correlates and Predictors of Growth or Decline

Structural equations modeling provides researchers with a means for specifying and modifying theoretical models for promoting theory development and construct validation (e.g., Bentler, 1978; Huba & Bentler, 1982; Raykov, Tomer, & Nesselroade, 1991). This methodology permits the development of unambiguous models that represent important theoretical frameworks and which allow the examination of the influence of latent “error-free” constructs on other constructs, all measured with fallible indicators.

Within the framework of the SEQM methodology, a general model for measuring residualized change and discerning patterns of ability growth or decline can be described as follows. Assume that at each of two measurement points under consideration, several congeneric measures are available as indicators of the repeatedly measured psychological construct and of the latent dimension measured by the presumed/studied change correlates. Congeneric tests measure the same latent attribute in possibly different units of measurement, scale origins, and degree of precision (e.g., Jöreskog & Sörbom, 1988). These measures are denoted by X'_1, X'_2, \dots, X'_k ; $X''_1, X''_2, \dots, X''_l$; and $X_{c1}, X_{c2}, \dots, X_{cm}$ for the pretest, posttest, and correlates, respectively (see Raykov, 1992a, 1992b for two special cases in the form of equivalent models), where k and l may differ from one another or from m . Either k or l should be greater than 1, because otherwise the model will be underidentified unless further—generally unreasonable—restrictions are introduced. (The model below does not impose the constraint that the same measures be used at the two assessments, as long as the sets of k and l congeneric tests assess the same psychological construct. That is, if required, only a part of the congeneric measures used at pretest or posttest may be administered at posttest or pretest, provided they measure the same repeatedly assessed latent ability dimension.)

Extending the underlying idea of the Tucker et al. (1966) approach, the model definition equations are:

$$X'_1 = a'_1 T' + E'_1, X'_2 = a'_2 T' + E'_2, \dots, X'_k = a'_k T' + E'_k \quad (2)$$

for the pretest;

$$X''_1 = a''_1 T'' + E''_1, X''_2 = a''_2 T'' + E''_2, \dots, X''_l = a''_l T'' + E''_l \quad (3)$$

for the posttest;

$$X_{c1} = a'''_1 T_c + E_{c1}, X_{c2} = a'''_2 T_c + E_{c2}, \dots, X_{cm} = a'''_m T_c + E_{cm} \quad (4)$$

for the correlates; and

$$T'' = bT' + d + g \quad (5)$$

$$\text{Cov}(T', g) = 0. \quad (6)$$

In this system of $k + l + m$ equations, errors of measurement in the observed variables are signified by E . The constants a'_i , a''_j , and a'''_p ($i = 1, \dots, k$; $j = 1, \dots, l$; and $p = 1, \dots, m$) are unknown model parameters, as are the error variances, the variance of the residualized gain score (g), the

covariance between T' and T_c , the critical covariance between g and T_c (see below), the structural regression slope b , and the intercept d . The latter constant d in this model allows empirical situations to be investigated, such as those in which the underlying latent dimension changes by a (average) positive or negative shift, and is a substantively meaningful model parameter when analyzing the mean structure of the observed variables, because it reflects that part of the posttest mean that remains after partialing out its linear multiplicative prediction from the initial mean (see below).

In this model, Equation 5 is the basis of the study of correlates and patterns of true change. To achieve model identification, Equation 6 is included (e.g., Jöreskog & Sörbom, 1988). Also, for each of the three true score latent variables involved, either the variance or a regression coefficient of an observed variable on them should be set equal to a constant—for example, 1. [Fixing variances of latent dependent variables can be implemented as the default option in the recently developed structural modeling program RAMONA (see Browne & Mels, 1990, and below)].

Equations 2–6 show that this general model adopts the degree of covariation between the residualized measure of true change g and the studied correlates as a model parameter, which is accomplished by including the former as a separate latent independent variable. This variable,

$$g = T'' - bT' - d, \quad (7)$$

represents the structural residual remaining after partialing out that part of true posttest that is linearly predictable from true initial status. This residual's indices of interrelationship with other psychological dimensions are of major interest in studies aimed at identifying correlates and predictors of change patterns that are atypical for a group of persons.

The inclusion of the residualized true change g is achieved through modeling of the structural relationship between posttest and pretest in terms of Equation 5. That equation is incorporated into the model by setting the path from g to true posttest T'' equal to 1 and treating the structural intercept d as the latent mean of T'' when fitting the model under consideration to the empirical covariance/mean matrix (e.g., Bentler, 1989; Jöreskog & Sörbom, 1988).

By explicitly accounting for measurement error in the observed variables, this model yields an estimate of the relationship of residualized true gain with the correlate variables, which is free from the artifactual “part” of the dependence of observed change on initial status. That “part” results from the presence of measurement error in the pretest (e.g., Bereiter, 1963; Wilder, 1965). By considering at least two measures of the psychological construct under investigation and/or for measuring the latent dimension of the correlates, the present approach, in contrast to the traditional methods of change measurement, allows for a consistent and efficient multiple-variable-based estimation of the degree of covariation between the residualized pretest-to-posttest true change with the correlates. In multiple-group designs, the model permits the study of group differences or invariance with respect to residualized true change correlations with correlates or predictors of change (see below).

A special case of the model in which $T_c = X_{c1}$ ($E_{c1} = 0$) and there are no further indicators of the latent dimension of the correlate is applicable to situations in which the studied correlate/predictor of change is an “error-free” measure (see Raykov, 1992a). The approach is also useful in cases in which there is only one fallible indicator of the latent dimension of the correlate, if reasonable assumptions concerning its measurement error variance can be made in order to achieve model identification. For example, that error variance can be related to the error variance of another observed variable(s), or the former variance can be set equal to the variance of the observed correlate variable multiplied by $1 - r_{cc}$, where r_{cc} is the estimated reliability coefficient of the indicator of the latent criterion dimension (e.g., Bollen, 1989; Jöreskog & Sörbom, 1988).

Parameter Estimation and Model Testing

With regard to the implied covariance/mean matrix of the observed variables, as well as to the statistical estimation and model testing procedure implemented in structural modeling programs, such as LISREL (Jöreskog & Sörbom, 1988), EQS (Bentler, 1989), COSAN (Fraser, 1992), LISCOMP (Muthén, 1987), RAMONA (Browne & Mels, 1990), EzPATH (Steiger, 1989), SAS PROC CALIS (SAS Institute, 1990), this SEQM can be considered a special case of a confirmatory factor analytic model. Let $\mathbf{X} = (X'_1, \dots, X'_k; X''_1, \dots, X''_l; X_{c1}, \dots, X_{cm})^T$ be the vector of observed variables (where the superscript "T" denotes a transpose). Further, let

$$\mathbf{A} = \begin{vmatrix} a'_{11}, \dots, a'_{1k}, & 0, \dots, 0, & 0, \dots, 0 \\ 0, \dots, 0, & a''_{11}, \dots, a''_{1l}, & 0, \dots, 0 \\ 0, \dots, 0, & 0, \dots, 0, & a'''_{11}, \dots, a'''_{1m} \end{vmatrix}^T \quad (8)$$

be the matrix of regression coefficients of observed variables on corresponding true scores; these coefficients are model parameters. Finally, let $\mathbf{E} = (E'_1, \dots, E'_k; E''_1, \dots, E''_l; E_{c1}, \dots, E_{cm})^T$ be the vector of measurement errors, and $\mathbf{T} = (T', T'', T_c)^T$ be the vector of true scores, with

$$T'' = bT' + d + g. \quad (9)$$

Then, the initial $k + l + m$ model definition Equations 2–6 can be rewritten in the form of the general linear model:

$$\mathbf{X} = \mathbf{AT} + \mathbf{E}. \quad (10)$$

The following expressions for the covariance matrix Σ_x and the mean vector $\mathbf{M}(X)$ of the $k + l + m$ observed variables are obtained from Equation 10:

$$\Sigma_x = \mathbf{A}\Sigma_T\mathbf{A}^T + \Sigma_E \quad (11)$$

and

$$\mathbf{M}(X) = \mathbf{AM}(T), \quad (12)$$

where Σ_T , Σ_E , and $\mathbf{M}(T) = [\mathbf{M}(T'), b\mathbf{M}(T') + d, \mathbf{M}(T_c)]^T$ are the covariance matrices of true scores and measurement errors, and the mean true score vector, respectively. Equation 12 indicates that observed variable means are modeled in terms of true score means and regression coefficients of observed variables on true scores. The matrix Σ_E can be diagonal, and contains the variances of the measurement errors, which are free parameters in the model. The true score covariance matrix Σ_T is symmetric and is parameterized as follows (only its upper triangle is given here):

$$\Sigma_T = \begin{vmatrix} \pi_{11} & b\pi_{11} & \pi_{1c} \\ & b^2\pi_{11} + \pi_{gg} & b^2\pi_{1c} + \pi_{gc} \\ & & \pi_{cc} \end{vmatrix}, \quad (13)$$

where

π_{11} is the variance of true initial status, T' ,

π_{gg} is the variance of the residualized gain score, g ,

π_{cc} denotes the variance of the true score T_c associated with the correlates, and

π_{1c} and π_{gc} designate the covariances between T_c and T' , and g and T_c , respectively.

Equations 11, 12, and 13 show that the model under consideration yields exactly the same parameterization of the observed variable covariance/mean matrix that would be obtained with a confirmatory factor analytic approach. In other words, when using SEQM software for fitting the model, its differences from the corresponding factor analytic model—which is a more general model because

the observed variable residuals also can contain variable specificity not accounted for by the corresponding common factor—are immaterial.

Compared to the residualized difference score or the Tucker et al. (1966) approach, when fitting the model under consideration to the analyzed matrix of observed variable interrelationships, the corresponding testing procedure tests whether the consequences of all model assumptions are inconsistent with the data, as represented by that empirical matrix. A proper application of this general SEQM to studying patterns, correlates, and predictors of change requires that it be tenable for data description and explanation. If the model is rejected when tested against a dataset, it may well yield misleading estimates of the critical correlation between residualized true change and correlates of growth/decline, or indicate an incorrect pattern of change.

Metric of True Change

Studying true change implies that a specific metric is selected as a frame of reference within which growth is assessed. This metric can be selected two ways—by fixing the variances of the latent variables to a constant (usually 1), or by fixing a path from each latent variable to one of its indicators (e.g., Bentler, 1989; Jöreskog & Sörbom, 1988). When using the described model for studying patterns of true change (see below), it is recommended that the first option be avoided, because fixing the variances of the true score latent variables to unity (or another constant) introduces a specific standardization of the latent metric. As a result, regression to the mean on the latent trait is artificially introduced, because then

$$\text{Var}(T'') = b^2\text{Var}(T') + \text{Var}(g) > b^2, \quad (14)$$

where $\text{Var}(T') = 1$, that is, $|b| < 1$ will hold, where $\text{Var}(\cdot)$ denotes variance (see below).

The specific research question and substantive considerations influence which congeneric measure should have a fixed regression coefficient at each assessment. The model then focuses on true change in terms of the specific units of measurement inherent in the selected measure. True change will be assessed, however, in the context of the other congeneric measures. The model will then provide a multiple-variable-based parameter estimation and test by taking into account information from measures related to the construct.

Selecting another repeatedly administered measure to be the indicator with fixed regression coefficients on the corresponding true scores T' and T'' will not change the overall goodness-of-fit indices of the model (the χ^2 and the descriptive goodness-of-fit indices). Also, fixing the regression coefficient to 1 of another measure on these latent variables will not change the estimate of the critical correlation between residualized true change and the latent dimension indicated by the correlates. This is because the model is scale-invariant, as is the maximum-likelihood fit function, and because the correlation coefficient is a scale-free model parameter (e.g., Cudeck, 1989).

However, the estimate of the structural regression slope and intercept parameters b and d , as well as the estimates of the regression coefficients of observed variables on corresponding true scores, will generally be changed because all these parameters are not scale-free, as their value depends on the units of measurement inherent in the underlying latent metric. Further, as mentioned above, to ensure model identifiability either one of the parameters $a_1''', a_2''', \dots, a_m'''$ or the variance of the covariates' true scores T_c has to be fixed to 1. Again, selecting any of the latter $m + 1$ options has no bearing on the overall goodness-of-fit indices, or on the value of the correlation of interest here $[\rho(T_c, g)]$. Because this correlation coefficient is scale-free, its estimate will not depend on whether a regression coefficient from an observed correlate variable on its corresponding true score variable or latent variances are fixed to 1 (the correlation coefficient is invariant under linear transformations underlying this change of scales).

Structural Equations Modeling Approaches to the Study of Change

The described general SEQM allows the study of change in a repeatedly followed psychological construct that is measured at each of several assessments with a number of congeneric indicators. The model is easy to implement, test, and estimate with existing SEQM software. It parameterizes the degree of interrelationship between residualized true change and correlates of growth/decline in the studied latent dimension, and the pattern of true change (see below). This model is related to other approaches to the measurement of change within the SEQM framework, which have been proposed recently (see Browne & Du Toit, 1991; McArdle & Anderson, 1990; Meredith, 1991; Muthén, 1991; Nesselroade, 1991; Tisak & Meredith, 1990a, 1990b).

The SEQM approach of McArdle & Anderson (1990; McArdle & Aber, 1990) models longitudinal change in single, possibly interrelated, repeatedly assessed psychological variables, without integrating them into latent variables at each measurement occasion. The latter integration is implemented within the model described here by employing several congeneric measures at each of a number of repeated measurements. By using the latter model with several congeneric measures for the first and last of a number of consecutive measurements, the estimation of the degree of interrelationship between residualized true overall change in the construct under investigation and the studied correlates is possible. Using a modification—to accommodate the covariates and their latent dimension—of the model proposed by McArdle & Anderson (1990) allows estimation of the degree of interrelationship between (unresidualized) true overall change in a repeatedly assessed psychological variable and correlates of change. Employing the multivariate variant of this modification permits interrelating, rather than integrating, the true overall change in each of the longitudinally followed variables and estimating the extent of covariation of this change in each variable with the correlates. However, even though this integration is achieved within the model described here, in contrast to that of McArdle & Anderson (1990) the present method does not permit testing hypotheses regarding specific patterns of change in a repeatedly assessed variable(s) across several measurements. A selected hypothesis—for example, linear growth—can be tested directly with the McArdle & Anderson (1990) model.

Tisak & Meredith (1989, 1990a, 1990b) proposed a comprehensive exploratory factor analytic approach to the study of change. Their general longitudinal factor analysis model encompasses the measurement part of the model described here. In contrast to the present model, the Tisak & Meredith (1990a) method does not model the structural relationship between the factors representing the repeatedly followed latent dimension.

Muthén (1991) developed a random effects growth model that estimates parameters reflecting individual growth as well as parameters reflecting the group growth pattern. Specifically, this model permits the introduction of individual intercepts as parameters. In contrast to this approach, the model presented here can be considered a conventional fixed effect SEQM-based technique that allows for a certain parsimony with regard to model parameters and addresses the study of correlates and predictors of residualized growth/decline in repeated assessments, as well as patterns of change.

Browne & Du Toit (1991; Harlow, 1991) considered a first-order autoregressive moving average time series model, a stochastic parameter growth curve model, and a latent growth curve model. Their approach can be extended to encompass more than one covariate, but currently it is mainly limited to the single-indicator case. That is, when applying their very general method, it is important to have very reliable indicators of the underlying latent dimension of interest. The models of Browne & Du Toit (1991) are useful when there are at least three repeated measurements (preferably more than five; see Harlow, 1991) in order to provide stable parameter estimates and to reliably model the pattern of growth. In contrast to their models, the method described here is specifically designed

for two assessments with multiple indicators. However, Browne & Du Toit's (1991) models represent a unified general approach that takes into account intraindividual as well as interindividual information, which provides for a rigorous evaluation of patterns of latent growth/decline, including curvilinear relationships.

Meredith (1991; McArdle, 1991) focused on change factors. His latent change and difference analysis model has relatively few parameters, as compared to conventional repeated measures analysis-of-variance approaches. In contrast to the method presented here, which builds on modeling the structural relationship between the repeated assessments of the psychological construct under study, Meredith's approach is based on modeling means of observed and latent variables, and is not specifically designed to include correlates of change.

Discerning Patterns of True Change

Within the model, the structural regression slope parameter b reflects the extent to which regression to the mean, overcrossing, or fan-spreading are in effect on the latent trait. Regression to the mean is observed when the mean of all posttest scores X'' , which follow a pretest score x' , is closer to the posttest mean $M(X'')$ than a particular pretest x' is to the pretest mean $M(X')$, where M is the expectation with respect to individuals. In other words, this phenomenon is present when a tendency is observed for scores above or below the mean at an initial measurement to fall closer to the mean at a subsequent assessment.

In formal terms (see Nesselroade et al., 1980; Rogosa & Willett, 1985; Wilder, 1965), regression to the mean is observed in the population when for the posttest expectation, conditioned on pretest x' , $M(X''|x')$, the following inequality in terms of absolute values holds true:

$$|M(X''|X' = x') - M(X'')| < |x' - M(X')| . \quad (15)$$

In this inequality, the original metric of the scores is preserved. For standardized scores, regression to the mean is observed in the recorded values if the mean of all posttest X'' following a particular pretest x' is closer to $M(X'')$ than to $M(X')$, when distances are measured in standard deviation units (e.g., Rogosa & Willett, 1985). This relation is satisfied whenever the correlation between pretest and posttest is less than 1, regardless of whether this imperfection is due to measurement error or to heterogeneity in individual (true) change (e.g., Healey & Goldstein, 1978). Thus, in terms of standardized scores, regression to the mean can be regarded as "a harmless mathematical tautology, and one which provides little insight for the study of change" (Rogosa & Willett, 1985, p. 217). Therefore, only unstandardized scores are used here.

From the definition of the model (Equation 5), the following equation is obtained for true posttest expectation conditioned on true pretest, when the model is tenable (see Equation 15):

$$M(T''|T' = t') = bt' + d . \quad (16)$$

Hence, for the variance of this conditioned expectation (which is a function of t') the following equality holds true (t' is any value in the range of the true pretest):

$$\text{Var}[M(T''|T')] = b^2\text{Var}(T') . \quad (17)$$

From this equation, it is clear that the structural regression parameter b reflects the relationship between average true pretest distance from the initial measurement mean and the average distance of the conditioned (on true initial status) true posttest to the mean of the posttest. For the latter mean, $M[M(T''|T')] = M(T'')$ holds true (e.g., Rao, 1965).

Next, from the model definition in Equation 5, and Equation 16, the following is derived:

$$M(T'' | T' = t') - M(T'') = bt' - bM(T') = b[t' - M(T')] \quad (18)$$

or, in terms of absolute values:

$$|M(T'' | T' = t') - M(T'')| = |b| |t' - M(T')|. \quad (19)$$

Thus, within the model under consideration, b represents a proportionality constant that equals the ratio between conditioned posttest distance from the posttest mean and pretest distance from the initial mean. If the model under consideration is not rejected by the data when its consequences for the implied mean structure are tested against the empirical covariance/mean matrix, interesting conclusions concerning the presence or absence of regression to the mean, overcrossing, or fan-spreading on the latent trait can be drawn on the basis of the confidence interval for b .

Figure 1 illustrates some of the kinds of relationships between true pretest variance and variance of the conditioned (on true pretest) posttest scores that can be discerned by the model, for cases in which means do not change across the two assessments. In Figure 1, the lines connect true scores at pretest [MP(1)] with corresponding expected true posttest scores [conditioned on true initial level; MP(2)]; $d = 0$ was assumed for all cases.

Regression to the Mean

Equation 19, in conjunction with Equation 15, shows that regression to the mean is in effect on true scores if and only if $-1 < b < 1$ holds true. Thus, testing for regression to the mean with respect to true scores is a test of the null hypothesis $H_0: -1 < b < 1$ against the composite alternative $H_1: b > 1$ or $b < -1$. Furthermore, according to Equation 19, the smaller $|b|$ is (i.e., the closer b is to 0), the more severe is the effect of regression to the mean on ability scores. Hence, in order to test for regression to the mean with respect to true scores, determine whether this interval has common points with the interval $(-1, 1)$. If this is not the case (Figures 1b, 1g), there is no evidence for regression to the mean with respect to true scores (see subcases B1 and B2 below). If the confidence interval of b has common points with the interval $(-1, 1)$, the null hypothesis of regression to the mean with respect to true scores cannot be rejected (e.g., Figures 1c and 1e).

If the null hypothesis of regression to the mean with respect to true scores is not rejected, the following three subcases are of further interest:

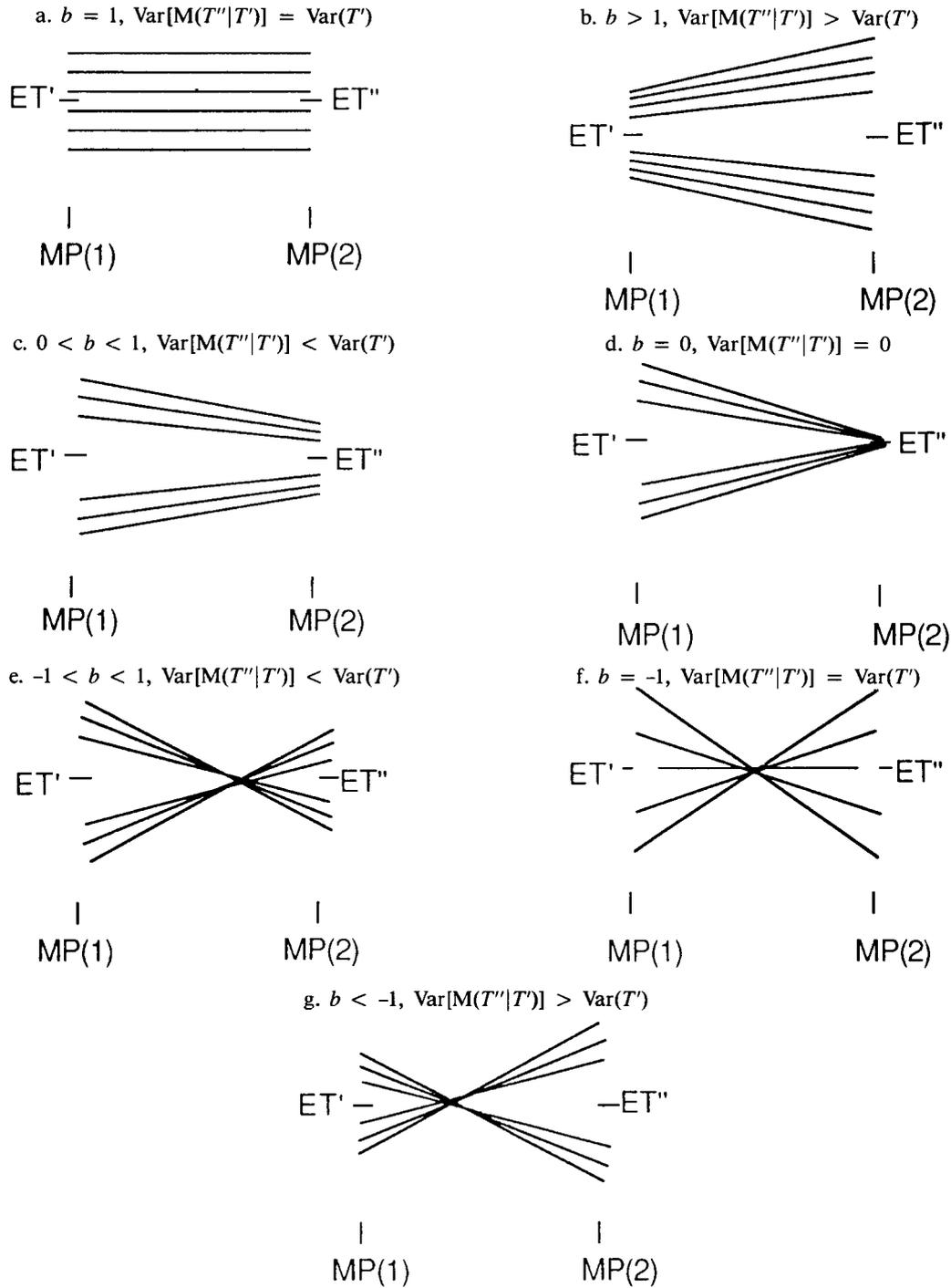
Subcase A1. If the confidence interval for b lies completely within the interval $(-1, 0)$, it suggests an ("mild," compare to subcase B1 below) overcrossing pattern of true change. True scores that are above or below the mean at the pretest are followed, on the average, by true posttest scores that are below or above the mean at the posttest. The latter are closer—in terms of absolute distances—to their mean than are the former to the pretest mean (see Equation 19 and Figure 1e). The severity of this overcrossing is reflected in b : The smaller b is (i.e., the closer b is to -1), the more severe is this overcrossing of true scores.

Subcase A2. If the confidence interval for b lies completely within the interval $(0, 1)$, it suggests that true scores above or below the mean at the pretest are followed, on the average, by true scores that are above or below the mean at the posttest. The latter are closer to their mean than are the former to the pretest mean (see Equation 19 and Figure 1c). The severity of this phenomenon is again reflected in b : The smaller b is (i.e., the closer b is to 0), the more severe is this form of regression to the mean with respect to true scores.

Subcase A3. If the confidence interval for b contains the point (-1) and lies completely below 0 (see Equation 19 and Figures 1e, 1f, and 1g), there is indication of an overcrossing pattern of change in true scores. True pretest scores above or below the mean are followed then, on the average, by

Figure 1

Relationship Between True Pretest Variance and Variance of the Conditioned (on True Pretest) Posttest



true posttest scores that fall below or above the mean. However, in this subcase, more precise statements concerning the severity of overcrossing with regard to (absolute) distances from the mean, such as those in the two previous subcases, cannot be made with this method. Similarly, if the confidence interval for b contains the point $+1$ and lies completely above 0 , then there is evidence for no overcrossing of true scores.

Each of the specific hypotheses corresponding to Figures 1a, 1d, and 1f also can be tested. The constraints $b = 0$, $b = 1$, or $b = -1$, respectively, can be imposed and the significance of the difference in the χ^2 values of the resulting two nested models can be determined (e.g., Jöreskog & Sörbom, 1988).

Egression From the Mean

If the confidence interval for the slope parameter b has no common points within the interval $(-1, +1)$, it is interpreted as “egression” from the mean with respect to true scores (Nesselroade et al., 1980; see Figures 1b and 1g). [The confidence interval for b cannot contain -1 and/or $+1$. Here, confidence intervals are used as open intervals by definition. Probabilistically, because maximum likelihood estimates are (continuously) asymptotically normally distributed, the probability that any of the ends of this interval is $+1$ or -1 is 0 ; see Rao, 1965.] Such a result would indicate an average divergence—in absolute values—of true scores from their mean, which occurs at posttest, as compared to the corresponding distance at pretest. In other words, at the second measurement the absolute value of the difference between the expectation of true scores conditioned on the pretest and the mean would be larger than the absolute value of the distance between the corresponding true score to the mean of the first measurement (see Equation 19). b indicates the extent of this (average) egression: The larger $|b|$ is than 1 , the more pronounced is this pattern of change. Two subcases can be considered.

Subcase B1. If the confidence interval for b lies completely below -1 (see Equation 19 and Figure 1g), it indicates an overcrossing pattern of true change. In this case, true pretest scores above or below the mean of the initial measurement are followed, on the average, by true posttest scores that are below or above the posttest mean. In contrast to Subcase A1, however, true posttest scores here tend to be further apart (in terms of absolute values) from their mean than their corresponding true pretest scores are from the mean of the first measure. The severity of this overcrossing effect is measured by b : The larger $|b|$ is (i.e., the smaller b is than -1), the greater is the severity of overcrossing.

Subcase B2. If the confidence interval for b lies completely above $+1$ (see Equation 19 and Figure 1b), it suggests egression from the mean with no overcrossing, that is, fan-spreading. True pretest scores above or below the mean pretest score are followed, on the average, by true posttest scores that are above or below the mean of the posttest scores. In contrast to Subcase A2, however, true posttest scores tend to be further apart—in terms of absolute values—from their mean than their corresponding true pretest scores are from the pretest mean. The severity of this fan-spreading effect is measured by b : The larger $|b|$ is (i.e., the larger b is than $+1$), the greater is the severity of fan-spreading. Thus, in multisample analyses, the model described here allows for studying group differences or invariance with respect to patterns of true change.

Illustration With Data From a Cognitive Intervention Study of Elderly People

Data

The data were from a two-group cognitive intervention study of plasticity in fluid intelligence of 248 healthy elderly people (Baltes et al., 1986). The experimental and control groups consisted of

161 and 87 persons, respectively. A battery of eight tests was administered to all persons as a pretest. Three posttests were administered after the experimental group underwent an intervention. The repeated assessments took place one week, one month, and six months after the training. The intervention comprised ten one-hour sessions of cognitive training involving the fluid intelligence subabilities induction (inductive reasoning ability) and figural relations. A detailed description of all measures and particulars about the training program are provided in Baltes et al. (1986), Baltes & Willis (1982), and Blieszener, Willis, & Baltes (1981).

For illustrative purposes, four measures of fluid intelligence are discussed—the ADEPT Induction (AI), ADEPT Figural Relations (AF), Thurstone's Standard Induction (TI), and Culture-Fair (CF) tests. The SEQM described above was used to study correlates and the pattern of growth in inductive reasoning ability, with the fluid intelligence subability figural relations used as a studied correlate latent dimension. In Baltes et al. (1986), the induction factor was marked by AI and TI, whereas figural relations was indicated by AF and CF. Hence, when applying that model, $k = l = m = 2$ was substituted in its definition (see Equations 2–6). To compare the experimental and control groups on a common scale, the pretest results on AF and CF were considered markers of the studied correlate latent dimension—figural relations. (At pretest the two groups did not differ with respect to any measure used.)

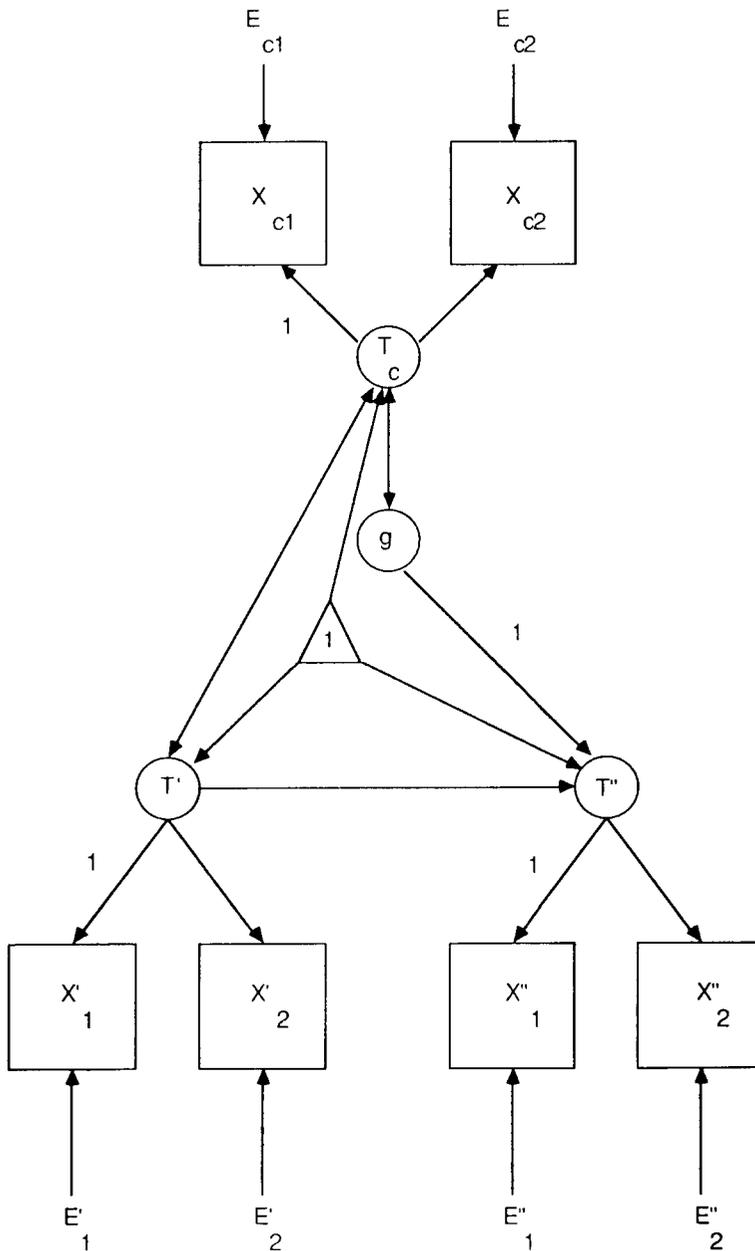
The SEQM Model

True change was studied in the metric of the AI. This typical inductive reasoning measure involved letter series, number sets, and letter sets, and was specifically constructed for the purposes of the study (e.g., Baltes et al., 1986; Baltes & Willis, 1982; Blieszener et al., 1981). It was therefore interesting to simultaneously study change patterns and correlates of induction growth that were residualized from the within-group pattern of change in inductive reasoning, as reflected in the measurement units inherent in the AI. Hence, this test was selected to be the measure with the path from true pretest and posttest fixed to 1 at the two assessments. That is, $a'_1 = a''_1 = 1$ was substituted in Equations 2–6.

The two-dimensional version of the general SEQM described above is graphically represented in Figure 2, which uses conventional path analytic graphic notation (e.g., Bentler, 1989; Jöreskog & Sörbom, 1988; Raykov et al., 1991). In the figure, as well as Equations 2–6, X'_1 and X'_2 represent the pretests for AI and TI, and T' is the pretest inductive reasoning ability score. X''_1 and X''_2 are the posttests for these measures, and T'' is the corresponding inductive reasoning ability score. X_{c1} and X_{c2} denote the initial assessments with the AF and CF measures, and T_c is the initial figural relations score. The unit diamond in the center of Figure 2 has a technical role in the ensuing mean structure analysis and is used in the diagram based on Equation 12. [Its presence, however, is not necessary if the model is fitted to the empirical covariance/mean matrix using LISREL 7 (see Jöreskog & Sörbom, 1988)].

Formally, when fitting the model, g is correlated with the residual of T_c , because the latter variable is regressed on the unit constant (see Bentler, 1989; McArdle & Epstein, 1987). The parameters of major interest were: (1) the covariance/correlation between residualized true change g with the correlate true score T_c , which is symbolized in Figure 2 by the two-way arrow connecting them; (2) the structural regression slope b reflecting the pattern of induction growth, which is signified by the path from T' to T'' ; (3) the structural intercept d that reflects residualized mean growth—that is, the part of posttest mean remaining after that part of the latter mean is partialled out, which is linearly and multiplicatively predictable from the pretest mean—and is symbolized by the one-way arrow leading from the unit constant to T'' . To achieve model identification, the parameter a'''_1 also was fixed to 1. In essence, this model indicates that, for each group, initial figural relations ability was assumed

Figure 2
A Two-Dimensional Variant of the SEQM Used



to be correlated with residualized growth in inductive reasoning ability.

Baltes et al. (1986; see also Raykov, 1992c) found indications of a significant increase in mean performance on AI and TI, which was maintained across the six-month testing period. This increase was more salient in the experimental group and outperformed the one resulting from the practice/retest effects in the control group. To explore the question of comparability of the pattern of overall change in the latter group to that indicated in the experimental group immediately after the training (relative to starting position), the focus was on pretest and final posttest in the control group, and pretest and first posttest in the experimental group. By fitting the model proposed here for these data simultaneously in the two groups, it was possible to consistently and efficiently estimate, and test for group invariance in, the degree of interrelationship between figural relations and the corresponding residualized within-group growth in inductive reasoning ability. In addition, the underlying ability growth pattern could be compared across groups.

Analysis and Results

Covariances, variances, correlations, and means of the six variables for the two groups are given in Table 1. Because the estimates of the univariate skewness and kurtosis, as well as that of Mardia's coefficient of multivariate kurtosis (e.g., Bentler, 1989) did not indicate departure from normality, the maximum likelihood estimation and testing procedure was used. First, the model was fit to the data with the correlation between residualized induction growth and figural relations constrained for equality across groups. (This constraint was implemented by restricting the variances of g and T_c and their covariance for equality across groups.) This yielded an acceptable χ^2 value of 30.413 with 21 degrees of freedom (df ; $p = .08$). Releasing this group restriction within the next model variant, in which the former was nested, yielded a decrease in the χ^2 to 25.766 with 18 df . Because the 4.766 difference in χ^2 values was not significant, the null hypothesis that the correlation between residualized induction gain and figural relations was the same in the experimental and control groups was not rejected. The common correlation was estimated at .18, with a significantly larger than 0 corresponding covariance. This result suggested that there was a discernible group-invariant (linear)

Table 1
Variances (Main Diagonal), Covariances (Lower Triangle), Correlations
(Upper Triangle), and Means for AI, TI, AF, and CF on Repeated Occasions for
the Experimental and Control Groups [From Baltes, Dittmann-Kohli, & Kliegl (1986)]

Variable and Occasion	Variable and Occasion						Mean
	AI-1	TI-1	AI-2/4	TI-2/4	AF-1	CF-1	
Experimental Group ($N = 161$)							
AI-1	218.057	.853	.820	.796	.651	.681	30.826
TI-1	220.921	307.461	.869	.871	.719	.765	37.487
AI-2	219.379	276.086	328.112	.924	.726	.750	49.083
TI-2	228.180	296.517	324.969	377.212	.731	.737	53.299
AF-1	155.152	203.467	212.465	279.249	260.760	.761	49.483
CF-1	127.006	169.397	171.556	180.758	155.219	159.384	48.391
Control Group ($N = 87$)							
AI-1	217.993	.833	.801	.803	.626	.627	30.091
TI-1	219.873	319.222	.873	.896	.590	.659	38.616
AI-4	214.599	282.863	328.862	.909	.585	.702	43.790
TI-4	246.660	333.062	343.052	432.697	.634	.759	48.011
AF-1	149.861	171.006	171.889	213.759	262.786	.722	51.971
CF-1	117.886	149.819	161.992	200.823	148.860	161.981	49.812

relationship between initial figural relations and residualized induction growth. With regard to the control group, a possible explanation was to assume that those among them who showed larger than average growth in induction tended to be successful in using interrelationships between the fluid intelligence subabilities—figural relations and induction. Persons in the control group who exhibited a pattern of change below the group average induction growth, however, tended not to be able to profit to a comparable extent from the interrelations among these abilities.

Next, the question of whether there was an indication of regression to the mean on the latent induction level was addressed. To study this, the upper part of the model in Figure 2 containing the figural relation measures and their true score was deleted, and the remaining model was fitted to the covariance/mean matrix of the four remaining observed variables. (These variables were the repeated assessments with the AI and TI measures.) This model first was fit by imposing the cross-group parameter restriction of identity in the ability growth pattern—that is, restricting the structural regression slope b for group equality. The resulting χ^2 value was 6.923 with 7 df ($p = .44$). Releasing this cross-group constraint yielded a nonsignificant decrease in the χ^2 to 5.955 ($df = 6$). It was concluded that there was no indication of group differences with respect to the structural regression slope b that reflected patterns of change on the latent inductive reasoning ability level. In the former, restricted model variant, the estimate of this slope was 1.243 in the two groups, with a standard error of .043. The resulting 95% confidence interval was (1.159, 1.328). This interval indicated regression from the mean with respect to true scores in both groups (i.e., Subcase B2; see Figure 1b), rather than regression to the mean.

Because group differences in mean growth were reflected within the tenable model version in the structural intercept parameter d , the next fitted variant imposed the restriction of group equality in this parameter (in addition to the slope restriction). The resulting χ^2 was significantly higher than that obtained for the initial tenable model version ($\chi^2 = 18.504$, $df = 8$). Because the 11.581 difference in χ^2 values was significant, it was concluded that there was evidence of differences in mean growth, as reflected in this parameter. In the starting tenable model version, the estimate of this intercept parameter was significantly higher in the experimental group—here it was 11.129 with a standard error of 1.44, whereas in the control group it was 6.132 with a standard error of 1.591. This finding is consistent with previous mean growth-based analyses of the same data (e.g., Baltes et al., 1986; Raykov, 1992c).

The results in the second part of this analysis suggested that individual inductive reasoning ability scores tended to diverge from the mean to the same extent in the experimental and control groups. More specifically, with regard to induction growth in both groups, persons with high initial ability seemed to have exhibited growth that was more salient than the average group induction gain. Those with low initial ability, however, gained, on average, less than the mean group growth. This pattern of change in inductive reasoning ability appeared essentially the same in the two groups (even though mean growth in the induction tests was more salient in the experimental group).

Conclusions

This paper focused on a general structural equation modeling approach to the issue of measuring change and studying correlates of growth or decline and patterns of true change. The model defined by Equations 2–6 permits consistent and efficient estimation of parameters reflecting theoretically and empirically interesting aspects of change, such as indices of covariation between residualized growth or decline and other variables, as well as parameters that indicate the pattern of true change. When used for discerning growth patterns, the method tests corresponding parameter inequalities by comparing their confidence interval endpoints with specific numbers +1, 0, and -1.

Inequalities cannot be cleanly tested with current structural modeling methods. Indeed, the parameters can be estimated and the model can be tested with and without the inequality constraints—for instance, using COSAN, or LISREL after introducing dummy variables (e.g., Fraser, 1992; Rindskopf, 1983, 1984). However, there is no straightforward procedure (in contrast to the case of testing equality restrictions) for comparing the two model fits in order to gain insight into the relevance of the imposed inequalities, because the likelihood ratio underlying the model testing procedure in the case of inequality constraints is asymptotically distributed according to a less well-known distribution (e.g., Chernoff, 1954; Shapiro, 1985).

When research interest focuses on use of this SEQM for purposes of studying true change patterns rather than correlates of change, safer substantive conclusions will generally result if the correlates, their true scores, and their covariances with T' and g are not included in the model. This is because in empirical applications of the described “full” model in Equations 2–6 a statistically acceptable overall model fit can be achieved because of a close-to-perfect fit in the part of the model encompassing the correlates and their true scores, at the expense of that part comprising the repeated assessments, their true scores and the residualized true change g , with the model not holding in the latter part. Conversely, the “full” model can be statistically rejected because it has serious misspecifications in the part dealing with the correlates, while the part comprising the repeated assessments is tenable. Finally, because the confidence interval of the structural regression slope is asymptotic (e.g., Jöreskog & Sörbom, 1988), more precise (i.e., more trustworthy) statements regarding correlates of residualized growth or decline and true change patterns are arrived at with large samples.

References

- Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserve capacity of the elderly in aging-sensitive tests of fluid intelligence: Replication and extension. *Psychology and Aging, 1*, 172–177.
- Baltes, P. B., & Willis, S. L. (1982). Plasticity and enhancement of intellectual functioning in old age: Penn State's adult development and enrichment project (ADEPT). In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 353–389). New York: Plenum.
- Belsky, J. (1990). *The psychology of aging: Theory, research, and interventions*. Pacific Grove CA: Brooks/Cole.
- Bentler, P. M. (1978). The interdependence of theory, methodology, and empirical data: Causal analysis as an approach to construct validation. In D. B. Kandel (Ed.), *Longitudinal drug research* (pp. 267–302). New York: Wiley.
- Bentler, P. M. (1989). *EQS: Structural equation program manual*. Los Angeles: BMDP Statistical Software.
- Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3–20). Madison: University of Wisconsin.
- Blieszener, R., Willis, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology, 2*, 247–265.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bond, L. (1979). On the base-free measure of change. *Psychometrika, 44*, 351–355.
- Browne, M. W., & Du Toit, S. H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47–68). Washington DC: American Psychological Association.
- Browne, M. W., & Mels, G. (1990). *RAMONA PC: User's guide* [Computer program and manual]. Columbus: Department of Psychology, Ohio State University.
- Burr, J., & Nesselroade, J. R. (1990). Change measurement. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Principles and structuring change* (Vol. 1, pp. 3–34). San Diego CA: Academic Press.
- Cavanaugh, J. C. (1990). *Adult development and aging*. Belmont CA: Wadsworth.
- Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics, 25*, 573–578.
- Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics, 10*, 637–666.
- Cronbach, L. J., & Furby, L. (1970). How should we “measure” change—or should we? *Psychological Bulletin, 74*, 68–80.
- Cudeck, R. (1989). Analyzing correlation matrices

- using covariance structure models. *Psychological Bulletin*, 105, 317-327.
- Fraser, C. (1992). *COSAN: User's guide* [Computer program and manual]. Faculty of Education, University of New England, Armidale, NSW, Australia.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Harlow, L. L. (1991). Comments on "Models for learning data." In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 69-74). Washington DC: American Psychological Association.
- Healey, M. J. R., & Goldstein, H. (1978). Regression to the mean. *Annals of Human Biology*, 5, 277-280.
- Huba, G. J., & Bentler, P. M. (1982). On the usefulness of latent variable causal modeling in testing theories of naturally occurring events: A rejoinder to Martin. *Journal of Personality and Social Psychology*, 43, 604-611.
- Isaac, P. D. (1970). Linear regression, structural relations, and measurement error. *Psychological Bulletin*, 74, 213-218.
- Jamieson, J., & Howk, S. (1992). The law of initial values: A four factor theory. *International Journal of Psychophysiology*, 12, 53-61.
- Jin, P. (1992). Toward reconceptualization of the law of initial values. *Psychological Bulletin*, 111, 176-184.
- Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and its applications*. Chicago IL: SPSS Scientific Software.
- Kendall, M. G., & Stuart, A. S. (1962). *The advanced theory of statistics* (Vol. 2). London: Griffin.
- Lacey, J. I., & Lacey, B. C. (1962). The law of initial values in the longitudinal study of autonomic constitution. *Annals of the New York Academy of Sciences*, 28, 1257-1290.
- Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research*, 49, 121-150.
- Lord, F. M. (1960). Large-sample covariance analysis when the control variable is fallible. *Journal of the American Statistical Association*, 55, 307-321.
- Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 21-38). Madison: University of Wisconsin Press.
- Lord, F. M., & Novick, R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.
- Manning, W. H., & DuBois, P. H. (1962). Correlational methods in research on human learning. *Perceptual and Motor Skills*, 15, 287-321.
- McArdle, J. J. (1991). Comments on "Latent variable models for studying differences and change." In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 164-169). Washington DC: American Psychological Association.
- McArdle, J. J., & Aber, M. S. (1990). Patterns of change within latent variable structural equation models. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Principles and structuring change* (Vol. 1, pp. 151-224). San Diego CA: Academic Press.
- McArdle, J. J., & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 22-44). New York: Academic Press.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110-133.
- McCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, 100, 107-120.
- McCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490-504.
- Meredith, W. (1991). Latent variable models for studying differences and change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 149-163). Washington DC: American Psychological Association.
- Messick, S. (1981). Denoting the base-free measure of change. *Psychometrika*, 46, 215-217.
- Muthén, B. O. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model* [Computer program and manual]. Mooresville IN: Scientific Software.
- Muthén, B. O. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 1-17). Washington DC: American Psychological Association.
- Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 92-105). Washington DC: American Psychological Association.
- Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin*, 88, 622-637.
- Rao, C. R. (1965). *Linear statistical inference and its applications*. New York: Wiley.
- Raykov, T. (1992a). Base-free measurement of change: A structural equation modeling approach. *Zeitschrift fuer Psychologie*, 200, 79-86.
- Raykov, T. (1992b). Structural models for studying correlates and predictors of change. *Australian Journal of Psychology*, 44, 101-112.

- Raykov, T. (1992c). On structural models for analyzing change. *Scandinavian Journal of Psychology, 33*, 247-265.
- Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in psychology and aging: Some proposed guidelines. *Psychology and Aging, 6*, 499-503.
- Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models. *Psychometrika, 48*, 73-83.
- Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*, 37-47.
- Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90*, 726-748.
- Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Psychology, 20*, 335-343.
- Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228.
- SAS Institute. (1990). *SAS Technical Report P-200. SAS/STAT Software: CALIS and LOGISTIC Procedures. Release 6.04*. Cary NC: SAS Institute.
- Schaie, K. W., & Hertzog, C. (1985). Measurement in the psychology of adulthood and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 61-92). New York: Van Nostrand Reinhold.
- Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika, 72*, 133-144.
- Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH* [Computer program]. Evanston IL: SYSTAT.
- Tisak, J., & Meredith, W. (1989). Exploratory longitudinal factor analysis in multiple populations. *Psychometrika, 54*, 261-281.
- Tisak, J., & Meredith, W. (1990a). Longitudinal factor analysis. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Principles and structuring change* (Vol. 1, pp. 125-150). San Diego CA: Academic Press.
- Tisak, J., & Meredith, W. (1990b). Descriptive and associative developmental models. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Time series and categorical longitudinal data* (Vol. 2, pp. 387-406). San Diego CA: Academic Press.
- Tucker, L. R. (1979). Comment on a note on a base-free measure of change. *Psychometrika, 44*, 357.
- Tucker, L. R., Damarin, F., & Messick, S. (1966). A base-free measure of change. *Psychometrika, 31*, 457-473.
- Webster, H., & Bereiter, C. (1963). The relationship of changes measured by mental test scores. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 39-59). Madison: University of Wisconsin Press.
- Wilder, J. (1965). Pitfalls in the methodology of the law of initial value. *American Journal of Psychology, 19*, 577-584.
- Woodruff-Pak, D. S. (1988). *Psychology and aging*. Englewood Cliffs NJ: Prentice Hall.
- Zimmerman, D. W., & Williams, R. H. (1982a). The relative error magnitude in three measures of change. *Psychometrika, 47*, 141-147.
- Zimmerman, D. W., & Williams, R. H. (1982b). Gain scores in research can be highly reliable. *Journal of Educational Measurement, 19*, 149-153.

Acknowledgments

Parts of this research were supported by fellowships granted by the Max Planck Society for the Advancement of Science, Penn State's College of Health and Human Development, and the Stockholm Institute of Education. The author is grateful to P. B. Baltes, J. R. Nesselroade, G. W. Lindgren, and L. Stankov for their support and encouragement. The author is indebted to the editor and two anonymous referees for valuable comments and suggestions on a previous draft, and J. J. McArdle, P. M. Bentler, and M. W. Browne for valuable discussions on aspects of structural equations modeling and growth curve analysis. K. Bird, C. Stevens, and J. Polzehl provided useful comments on a version of the paper. The author is grateful to Paul B. Baltes, Freya Dittmann-Kohli, and Reinhold Kliegl for permission to use their data from the project "Aging and Plasticity of Fluid Intelligence." The second part of the article is an extensively rewritten variant of a paper presented at the Seventh European Meeting of the Psychometric Society, which was held in Trier, Germany, in July 1991.

Author's Address

Send requests for reprints or further information to Tenko Raykov, University of Melbourne, Department of Psychology, Parkville, VIC 3052, Australia.