One of the central issues debated among personality psychologists has been the degree to which personality changes throughout adulthood. Although some theorists have argued strongly that change in personality should be an expected attribute of normal, adult development (e.g., Erikson, 1950; Levinson, 1986), empirical investigations of this question have often reported little if any change in personality in adulthood (e.g., Conley, 1984; Costa & McCrae, 1988; Kelly, 1955; Siegler, George, & Okun, 1979). These findings have led to a growing acceptance of the view that personality in adulthood is characterized much more by stability than by change (Conley, 1985; Field & Milsap, 1991; Kogan, 1990; McCrae & Costa, 1990; see Costa & McCrae, chapter 2 in this volume for a comprehensive summary). However, uncritical adoption of this view may be somewhat premature, at least in part because of important methodological and conceptual factors that can preclude a direct interpretation of the results of existing research on change in personality.
The study of personality development demands an interest in the analysis and measurement of individual change, an area that has long perplexed behavioral researchers (Bryk & Raudenbush, 1987). One complication arises due to frequent ambiguities resulting from the specific terminology used, often rather imprecisely, by various investigators. For example, terms like *continuity* and stability have been used interchangeably to indicate the lack of certain types of change in personality characteristics over time (cf. Shanan, 1991). This lack of precision has especially important consequences for research on personality development because, as we illustrate below, different types of stability can be assessed only through certain types of data analyses.

**TYPES OF STABILITY**

Several authors (e.g., Caspi & Bern, 1990; Ozer & Gjerde, 1989) have noted that the term stability can take on different meanings when applied to personality research. One important distinction differentiates between absolute stability and relative stability. Absolute stability refers to a lack of change in the absolute level of the measured attribute(s) over time. Relative stability typically refers to the consistency of the rank order of individuals within a group on some individual-differences measure across time.

Assessing Absolute Stability

An examination of absolute stability requires the analysis of a set of repeated measures of a personality attribute obtained from individuals over time. The changes in each individual's score on those measures allow an assessment of the degree of absolute stability for each individual.

However, absolute stability has often been assessed not at this individual level, but at an aggregate level. That is, stability has been assessed by comparing the mean value for a group of individuals measured at one time with the mean for the same group measured at a later time (e.g., Costa & McCrae, 1978; Siegler et al., 1979; see Caspi & Bern, 1990, for a review of this issue). However, the finding that the mean level of some personality attribute does not change over time does not necessarily indicate that there is no change at the individual level. If individual changes were random and both positive and negative, they could be canceled out by the process of averaging, thereby resulting in no mean change being observed. Thus,
the examination of change in the average level of a measured personality attribute may be uninformative regarding the degree of change at an individual level.

Assessing Relative Stability

Whereas the evaluation of absolute stability has sometimes mistakenly been assessed with cross-sectional data (e.g., Costa et al., 1986), the assessment of relative stability requires the collection of longitudinal data. The retest correlation coefficient (i.e., correlation between values at two times) is frequently used as the measure of relative stability (Costa & McCrae, 1989). Research using this methodology to assess relative stability suggests that there is considerable stability for a variety of different personality measures (see Caspi & Bern, 1990, for a summary).

Reliability and Estimates of Relative Stability

There are several problems with the use of the retest correlation coefficient as an estimated stability coefficient. First, this correlation is only an estimate of the true stability coefficient, one that will be attenuated by any unreliability of measurement. The retest correlation coefficient can be corrected for reliability-related attenuation using an estimate of the reliability of the measurement instrument. Using this approach, near-perfect stability estimates have been reported on several personality attributes (e.g., Costa & McCrae, 1988).

However, one problem in the use of this method is determining an appropriate estimate of reliability. This problem is quite serious, because in some cases the formula to correct correlations for attenuation due to low reliability of measurement may result in estimated correlation coefficients that are greater than one (see Ghiselli, Campbell, Sr. Zedeck, 1981, p. 242, for one hypothetical example).

This fact makes two commonly used methods of estimating reliability problematic. One method depends on the use of reliability estimates supplied with published tests. This method is not recommended (e.g., Anastasi, 1988) because typically, the sample used to generate the reliability estimates is not comparable with that to be included in the current study. Of particular importance is the heterogeneity of the samples involved. When a reliability estimate is obtained from a homogenous sample, the range of scores on the test will be restricted, resulting in a necessarily lower estimate of the reliability than if a more heterogeneous sample were used. And, the lower the reliability estimates, the more likely it will be that the test–retest correlation (i.e., the stability index) will be greater than one when corrected for attenuation. A seemingly preferable method would be to estimate reliability from a separate sample taken from the same population as the sample that
is the focus of the research. However, this alternative clu
estimated correlation that is greater than one.

Another solution is to estimate reliability of the personality measure based on the sample of individuals. For which the stability-coefficient is to be calculated. Of course, if only two waves of data are available, the reliability estimate must be some form of internal-consistency measure, because the test–retest correlation (i.e., the Time 1–Time 2 correlation) is precisely the statistic of substantive interest.

Other Methods for the Analysis of Relative Stability

In instances in which the personality attribute has been measured on three or more occasions, or when there is more than one indicator of the personality attribute assessed at each of two times, several alternative procedures exist for estimating relative stability. Although there are some specific differences among these procedures, they are all based on the assumption that individual development follows a first-order autoregressive model. Such models assume that an individual's current status is dependent only on that individual's previous status plus some random component (for an accessible introduction to these models, see Kenny & Campbell, 1989). That is, these methods model a person's score at time + 1 by regressing the t + 1 score on that individual's score at time t. Thus, only the previous measure (i.e., the score at time t) is seen to directly influence the current score (i.e., the score at time t + 1). Further, these models have traditionally only allowed for linear relationships between the data over time.

Processes that follow first-order autoregressive models produce correlation matrices that have a simplex structure, that is, a structure characterized by a matrix in which the elements of the matrix decrease in magnitude as they increase in distance from the main diagonal, and in which the elements along the subdiagonals are approximately equal (Jureskog, 1970). It has been suggested that such models are appropriate in longitudinal studies in which the same variable is measured for the same individuals over several occasions (Jureskog & Srbom, 1986).

From this general perspective, Heise (1969) has developed a procedure for use in situations in which there are three waves of data. This procedure allows one to estimate the reliability of the measurement instrument (which is assumed to be equal across occasions of measurement) and correct retest correlations for attenuation using this estimate. Costa, McCrae, and Ardenberg (1980) adopted this procedure and reported stability coefficients greater than 0.80 for the scales of the Guilford-Zimmerman Temperament Survey over a 12-year period.

---

1 For an example of a simplex-like correlation matrix (although one that is not necessarily derived from an autoregressive model), see Table I, Sample I.
Based on a similar model, when there are multiple indicators of a given personality attribute obtained on at least two occasions of measurement, it is possible to fit a latent structure model to the data using techniques for covariance structure analysis (e.g., EQS, Bender, 1989; and LISREL, Jureskog & SOrbom, 1986). Relative stability is examined by estimating the correlations between the latent variables representing the personality attribute at each time of measurement.

RELATIVE STABILITY, ABSOLUTE STABILITY, AND THE RATE OF CHANGE

Although the above procedures can be used to provide improved estimates of relative stability, it is important to remember that they do not provide any information with regard to absolute stability. Because estimates of relative stability are necessarily based on correlation coefficients, they essentially discard any information regarding differences in absolute levels of the variable across time. In the simplest case, a perfect linear relationship that increases in time would show perfect relative stability (i.e., the retest correlation would be equal to one), but could show a great deal of absolute change, whether assessed at an individual or at a group level.

In a slightly more complex situation, consider a case in which individuals, on average, become more introverted over time. If those who are most introverted at the time of first measurement increase in introversion more (or at a faster race) than those who were initially the least introverted (i.e., if there is a positive correlation between initial degree of introversion and increase in introversion over time), then this situation would result in perfect relative stability, but also absolute change at both the individual and the group level.

In other cases, however, differences in the rate of change of a variable over time can result in less than perfect relative stability. If the situation described above is reversed, and it is assumed that those who start out the least introverted have the fastest rate of change (i.e., there is a negative correlation between initial level of introversion and rate of change), it is easy to imagine a situation in which those who start out at the lower end of the introversion scale (relative to others in the group) could catch up to, and perhaps even surpass, those at the upper end. This situation would indicate low relative stability (as well as low absolute stability).

This discussion suggests that change should be conceptualized somewhat differently than has historically been the case. In the past, change has been conceptualized as some type of discrete increment or decrement in behavior occurring between two occasions of measurement (Willett, 1988). As a result, the most frequently used measure of individual change was the difference score. However, a substantial and often confusing lit-
ature arose regarding this measure (e.g., Cronbach & Furby, 1970; Harris, 1963; Linn & Slinde, 1977). The predominantly unfavorable appraisal of the difference score, and the implicit implication that it is an unsuitable measure of individual change, gained widespread acceptance that resulted in both a decrease in the use of difference scores by many investigators and a corresponding reluctance of some journal editors to publish results that were based on analyses of difference scores (Schaie & Hertog, 1985).

More recently, however, something of a counter movement has appeared in the psychometric literature that reevaluates the issues related to assessing individual change and provides an important new perspective on many of the previously presented problems with measuring individual change (see, e.g., Rogosa, Brandt, & Zimowski, 1982; Willett, 1988). The foundation of this alternative perspective is the belief that individual change or growth should be viewed not as a discrete process (e.g., as change), but as a continuous process (e.g., as growth) that underlies development. The change perspective has been called "unnatural" (Willett, 1988, p. 347) and has been seen as the source of many of the problems related to the measurement of change or growth.

For example, it has been argued that the conceptualization of change as a discrete process has promoted the reliance on two-wave longitudinal data as the basis for assessing change. Such two-occasion designs are limited in that they do not provide adequate data for analyzing and comparing individual differences in change (Bryk & Weisberg, 1977; Nesselroade, Stigler, & Bakes, 1980; Rogosa et al., 1982). Furthermore, even when data have been obtained on more than two occasions, the incremental conception of change has resulted in researchers often dividing the data into pairs of waves for purposes of analysis (e.g., Calsyn & Kenny, 1977; Eisdorfer & Wilkie, 1973).

When development is conceptualized as a continuous process, on the other hand, models for individual growth become the basis for the description and analysis of change (Rogosa et al., 1982). More specifically, under this approach, multiple waves of longitudinal data (preferably more than two) are collected from a sample of individuals, and an explicit model of individual development is specified and fitted to these observations. The parameters from the collection of individual growth curves then become the focus for subsequent statistical analyses.

It is possible to select from a wide variety of mathematical models to represent individual status as a function of time. These models range from the simple (e.g., conceptualizing growth as a linear function of time) to the comparatively complex (e.g., assuming growth follows some form of exponential or logistic pattern over time), with the choice of the individual growth curve model being made on either theoretical or empirical grounds.
The preferred alternative is to select a model for individual growth so that the parameters from the individual growth curves are rationally interpretable (i.e., they make sense in terms of what is theoretically known about the processes that underlie development; Willett, 1988). This approach has been used extensively in research investigating physical growth (e.g., Tanner, 1988). The strength of this approach relates to the fact that the parameters from the individual growth curves have real interpretative value. However, in the social sciences, models are typically derived empirically, using the polynomial model of the lowest degree necessary to provide an adequate fit to the data (Bryk & Raudenbush, 1987; Willett, 1988). Such empirical curve fitting is often used when there is little knowledge regarding the underlying mechanisms of growth (Guire & Kowalski, 1979; Willett, 1988).

Viewing change as a continuous process, and bearing in mind our earlier comments about the ways that rate of change can affect relative stability, it is clear that neither a measure of relative stability nor a measure of absolute stability alone provides a complete description of the nature of change at the individual level. It is preferable to have some way of describing and assessing both levels of absolute change and individual differences in the rate and pattern of individual change. The examination of growth curves described above (and more fully explained below) provides such a method. The choice of the model to fit the growth curves represents an examination of or decision about the pattern of change. Further, in the linear case, the rate of change can be assessed by examination of the individual slopes of the growth curves. In nonlinear cases, the parameters of the specified functions can be interpreted to obtain a variety of types of information regarding the change.

GROWTH CURVE ANALYSES

As discussed earlier, the first step in the analysis of growth curves is the adoption of an appropriate model to describe the nature of the change. One of the most commonly adopted models is the linear growth curve. Under this model it is assumed that individual development is a linear function of time, which can be represented as follows:

\[ Y_i(t) = \beta_0 + \beta_1 t + e_i \]

Equation 1 simply provides the mathematical description of a straight line with respect to time (t) for a specific individual i. As such, \( Y_i \) represents the observed score for individual i at time t; the intercept \( \beta_0 \), represents the true level of \( Y \) for individual i at time 0; the slope \( \beta_1 \), represents a
growth parameter, that is, the true race of change in \( Y \) for individual \( i \); and the term \( \epsilon_i \) represents random error in the measurement of \( Y \) for individual at time \( t \).

Suppose, for example, introversion is measured for a sample of individuals at several different times. Individual linear growth curves can then be fit to these data by regressing the observed introversion scores on the times of measurement. The growth curve parameters, \( \gamma_c \) and \( \gamma_{\gamma_c} \), represent the intercept and slope, respectively, from these individual linear regressions. In terms of analyzing individual races of change, the important parameter is the regression slope, \( \gamma_{\gamma_c} \), that provides us with an idiographic measure of the rate of change in introversion over time. Those people for whom the estimated slope is zero can be said to exhibit absolute stability over time. On the other hand, those individuals with estimated nonzero slopes show change in introversion over time, with positive slopes indicating growth, and negative slopes indicating decline in the level of introversion.

Further, the mean of the slopes provides a summary measure of the average amount of change for the sample of individuals, with a mean slope of zero indicating absolute stability at the group level and mean slopes other than zero indicating change in the average level of introversion over time. Similarly, the variance of the slopes provides an index of interindividual differences in individual change. In instances in which every individual's race of change is the same (i.e., \( \gamma_{\gamma_c} \) is constant across individuals), the variance of the slopes will be zero. When, on the other hand, the variance of the slopes is greater than zero, there are differences between individuals in the rate of change over time.

When individual development is modeled using individual growth curves, the matter of assessing relative stability is not as straightforward as with other approaches (e.g., Kenny & Campbell, 1989). Although this might seem to be a shortcoming of the growth curve approach, it is, instead, a strength, because a richer description of change at the individual level is provided. Specifically, if individual change is assumed to follow a linear model, then relative stability can exist in one of two ways. First, if every individual demonstrates the same rate of change over time (or, conversely, no change over time), then the variance of the individual slopes will be zero. This is a situation in which one would observe perfect relative stability. However, relative stability is also possible when individuals show different rates of change (recall, for example, our discussion earlier, in which the most introverted individuals at Time 1 change at a faster rate than less
introverted people, resulting in high relative stability). In this situation the variance of the individual slopes will necessarily be greater than zero. In such cases, relative stability can be assessed by examining the correlation between the individual slopes and the individual intercepts (i.e., by examining the relationship between rate of change and initial position on the attribute being measured). If the absolute value of this correlation is one (or close to one), then the rank ordering of individuals for the variable under investigation is consistent over time, and, thus, the group of individuals demonstrates relative stability on the measured characteristic.

ASSESSMENT OF SYSTEMATIC INDIVIDUAL DIFFERENCES IN CHANGE

The use of the correlation between the estimates of the individual slopes and the individual intercepts as an index of relative stability represents an example of a general strategy that can be used with growth curve analyses in order to identify systematic individual differences in change. One of the advantages of adopting a growth curve perspective is that it provides a logical and tractable approach to dealing with a wide variety of questions that have traditionally been associated with correlates of change. This is an important advantage because many of the perceived problems with difference scores have been raised specifically in relation to their use as criterion variables in correlation research (e.g., Kessler, 1977).

Two main concerns have been raised with regard to the use of difference scores in this context. First, it has been argued that difference scores are intrinsically less reliable that their component scores (e.g., Linn Slinde, 1977). Along these lines, Rogosa et al. (1982) have shown that the reliability of the difference score is a function of three factors: (a) the precision of measurement for the component scores (i.e., the reliability of the Time 1 and Time 2 measures), (b) the length of time between the two measurements, and (c) the variability in the true individual changes. The most critical of these components is the variability of the true changes. In fact, when there are substantial between-individual differences in within-individual change, it has been demonstrated that the difference score can be more reliable than the component scores (Rogosa & Willett, 1983; Zimmerman (Si. Williams, 1982). Conversely, when there are only minor between-individual differences in within-individual change, the reliability of the difference score will necessarily be low (Rogosa et al., 1982). The crucial point is that when there are no substantial differences in true, between-individual change, then the observed difference score is an unreliable measure of between-individual change (Rogosa, 1988).

The second main criticism of the use of difference scores as criteria in correlational studies is that the correlation between the difference scores
and the initial status gives an advantage to individuals with certain initial scores (e.g., Bereier, 1963; O'Connor, 1972; Plewis, 1983). This seems to be an invalid criticism, however, because the observed difference score is an unbiased, albeit fallible, estimate of true individual change (Rogosa et al., 1982). Further, Willett (1988) has argued convincingly that there may indeed be a correlation between true individual change and true initial status. Consider, for example, a situation in which all individuals are changing linearly over time (although at different rates), and a personality attribute (e.g., introversion) is assessed at more than two times. Given equal Time 1 scores, individuals who exhibit rapid growth will necessarily have higher Time 2 scores than those who grow less rapidly. If these patterns of individual growth continue unchanged, there will be a positive relationship between Time 2 scores and change, whether change is defined as the total amount of change between Time 1 and Time 3 or between Time 2 and Time 3. Thus, in situations in which individuals are changing over time it would appear that individual change and individual status would necessarily be related. A correlation between initial status and change, therefore, may reflect a true relationship, rather than an artifact of the measurement procedure.

The correlation between the difference score and initial status can take on any value from —1.00 to 1.00 (Rogosa et al., 1982; Rogosa & Willett, 1985b). Rogosa and his colleagues (Rogosa et al., 1982; Rogosa & Willett, 1985b) have argued that the frequent finding of a negative correlation between observed initial status and observed change is, at least in part, a function of the negative bias that exists in using the observed correlation between initial status and change to estimate the correlation between true initial status and true change. This bias results from the fact that the measurement error associated with the observed initial score is present, with the opposite sign, in the observed difference score. Thus, in situations in which there is a positive correlation between true change and true initial status, it is possible for the correlation based on the observed quantities to be zero or negative (e.g., R. L. Thomdike, 1966). In addition, the common practice of standardizing a variable so that it has a constant variance over time restricts the possible values of the correlation between change and initial status so that this correlation must necessarily be negative (Rogosa et al., 1982). Thus, the standardization of scores over time and the frequent reliance on the observed correlation between change and initial score as the estimate of the true relationship have both contributed to the

The negative bias that arises in using the correlation between observed initial status and observed change to estimate the correlation between true initial status and true change has long been recognized (e.g., E. L. Thorndike, 1924), and several alternative procedures exist for dealing with this negative bias (Blomqvist, 1977; Thomson, 1924; Zieve, 194C). Unfortunately, however, these adjustment procedures have seldom been used in empirical investigations.
misperception char change and initial status are necessarily negatively related.

Several of the proposed alternatives to the difference score as a measure of individual change have been derived to explicitly obtain measures of change that are uncorrelated with initial status. A somewhat confusing variety of residual change scores has emerged in the psychometric literature (e.g., the residual change score, DuBois, 1957; and the base-free measure of change, Tucker, Damarin, & Messick, 1966) and have been adopted for use in numerous empirical investigations (e.g., Arenberg, 1978). Interestingly, the psychometric properties, especially the reliabilities, of the various residual change scores are not that different from those for the simple difference score (Williams, Zimmerman, & Mazagatti, 1987).

In addition, there are severe logical and interpretational problems associated with the use of the various residual change scores. Residual change scores attempt to determine what the observed or true change would have been for a specific individual if all of the individuals in the sample had had equal initial scores (Rogosa et al., 1982). But, as explained earlier, there may actually be a relationship between initial status and change. Thus, answering this question discards information about "some genuine and important change in the person" (Cronbach & Furby, 1970, p. 74). Given these interpretational ambiguities, and a lack of clear improvement in the psychometric properties, residual difference scores cannot be considered to be better measures of individual change than simple difference scores.

The growth curve perspective suggests that the shortcomings of the difference score are not a result of any inherent deficiency with the difference score as a measure of individual change per se. Rather, this perspective leads to the realization that these shortcomings arise from the limitations imposed by the paucity of information provided with the common two-wave longitudinal design; two-wave data provide limited information, and do not allow for a comparison of individual differences in the rate or pattern of change.

Adoption of the growth curve perspective suggests that the development of models for individual change based on growth curves is the appropriate focus for the assessment of individual change and that the estimation of these models necessarily requires multiple waves of data.

The growth curve approach allows the examination of a wide variety of questions that deal with determinants of between-individual differences in individual change. These questions are formulated in terms of a second mathematical model, one that relates between-individual differences in the parameters of the individual growth curves to individual characteristics. The assessment of relative stability by examining the correlation between the individual slopes and intercepts is just one example of this general
More specifically, individual growth parameters can be modeled as follows:

\[ v_i = Rw + R_kX_t + \ldots + R_{l,p-1}X_{l,1} + U_{k,i} \]  

(2)

where there are \( p - 1 \) measured individual characteristics \( X_{k,i} \), \( \beta_{k,i} \) represents the effect of \( X \) on the growth parameter \( r_{i,1} \), and \( U_{k,i} \) is random error.

STATISTICAL ESTIMATION OF GROWTH CURVES

The two models represented in Equations 1 and 2 can be combined into a single, two-level model in order to investigate a diverse and rich variety of research questions that deal with individual change and systematic determinants of individual change. Models of this kind have been investigated under a variety of names within the statistical literature, including random coefficient regression models (e.g., Rosenberg, 1973), multilevel models (e.g., Goldstein, 1989), and hierarchical linear models (e.g., Ster-nio, Weisberg, & Bryk, 1983). Further, several possible alternative procedures have been developed in order to estimate such multilevel models. These vary from the relatively straightforward procedures outlined by Willett (1988) for use in estimating simple versions of such models to the more comprehensive estimation procedures, with their related computer software, that can be used for estimating more complex versions of the above models (e.g., Bryk, Raudenbush, Seltzer, & Congdon, 1988; Longford, 1986; Rasbash, Prosser, & Goldstein, 1989). Regardless of the estimation procedure used, however, these procedures provide an important alternative for use with longitudinal data, an alternative that has heretofore been largely ignored in the study of personality change in adulthood.

SIMULATED DATA FOR ANALYTIC COMPARISON

In order to demonstrate the potential utility of addressing questions of stability in longitudinal data from a growth curve perspective, several artificial data sets were generated. With artificial data the relationships among the true scores can be specified in advance, and random errors of measurement are added to these true scores in order to simulate observed scores. One of the inherent advantages of using artificial data is that various statistical techniques can be applied to the observed scores in order to

---

1 Equation 2 can be altered to incorporate nonlinear relationships between the parameters of the individual growth curves and individual characteristics.
determine how well they capture the nature of the underlying relationships among the true scores. A brief description of the simulation procedures precedes the presentation of the results from both conventional analyses and growth curve analyses.

Four hypothetical populations of 5,000 individuals were created under the assumption that, over time, levels of introversion change in a linear fashion. The true individual slopes and true initial values were sampled from a multivariate normal distribution. The populations differed in terms of the degree of relative stability imposed in the data. That is, the correlation between the true rates of growth and the true initial scores varied, with the correlations set to 0.00, 0.30, 0.60, and 0.90 for Populations I to 4, respectively. In each of the populations, the initial values were set to have a mean of 100 and a variance of 400, whereas the slopes were distributed with a mean of 0 and a variance of 25. Further, four waves of data were simulated using each of the populations by setting the times of observation as Time 1 = 0, Time 2 = 1, Time 3 = 2, and Time 4 = 3, and obtaining true scores corresponding to each time of observation through application of the individual growth curve equations. Random errors of measurement were drawn independently from a normal distribution with a mean of 0 and a variance of 71 and added to each of the true scores in order to produce, for each population, a set of observed scores, with the reliability of the Time 1 scores being equivalent to 0.83.

Finally, a random sample of 300 was obtained from each of the four populations to serve as the data for the following analyses.

Traditional Analyses of Stability

The bases for more traditional analyses of stability in longitudinal data are between-wave summary statistics. Table I provides the means, standard deviations, estimated reliabilities, and the between-wave correlation matrices for each of the samples of 4-wave data. Inspection of the summary statistics reveals several of the common attributes of longitudinal data, including increased variability with time (e.g., Krauss, 1980) and, with the exception of Table Id, the simplex-like correlation matrix in which the correlations decrease with increased distance from the main diagonal.

A common method of addressing questions related to absolute stability of mean level over time is to subject longitudinal data of the kind simulated...
### TABLE 1
Between-Wave Summary Statistics for Four Samples of Hypothetical Growth Data

(a) Sample 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.830</td>
<td>1.000</td>
<td>0.975</td>
<td>0.907</td>
</tr>
<tr>
<td>Time 3</td>
<td>0.775</td>
<td>0.845</td>
<td>1.000</td>
<td>0.996</td>
</tr>
<tr>
<td>Time 4</td>
<td>0.684</td>
<td>0.796</td>
<td>0.886</td>
<td>1.000</td>
</tr>
</tbody>
</table>

|     | 97.26 | 97.34 | 96.97 | 96.89 |
| M   |       |       |       |       |
| SD  | 21.05 | 22.09 | 24.19 | 26.80 |
| Reliability | 0.840 | 0.854 | 0.879 | 0.901 |

(b) Sample 2

<table>
<thead>
<tr>
<th>Time</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.833</td>
<td>1.000</td>
<td>0.981</td>
<td>0.961</td>
</tr>
<tr>
<td>Time 3</td>
<td>0.814</td>
<td>0.873</td>
<td>1.000</td>
<td>0.991</td>
</tr>
<tr>
<td>Time 4</td>
<td>0.776</td>
<td>0.863</td>
<td>0.903</td>
<td>1.000</td>
</tr>
</tbody>
</table>

|     | 99.13 | 97.60 | 98.45 | 99.51 |
| M   |       |       |       |       |
| SD  | 22.30 | 23.99 | 27.04 | 29.90 |
| Reliability | 0.857 | 0.877 | 0.903 | 0.920 |

(c) Sample 3

<table>
<thead>
<tr>
<th>Time</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.854</td>
<td>1.000</td>
<td>0.992</td>
<td>0.973</td>
</tr>
<tr>
<td>Time 3</td>
<td>0.827</td>
<td>0.882</td>
<td>1.000</td>
<td>0.995</td>
</tr>
<tr>
<td>Time 4</td>
<td>0.809</td>
<td>0.875</td>
<td>0.906</td>
<td>1.000</td>
</tr>
</tbody>
</table>

|     | 98.97 | 99.05 | 99.23 | 100.17 |
| M   |       |       |       |       |
| SD  | 20.68 | 24.20 | 26.83 | 30.15 |
| Reliability | 0.833 | 0.878 | 0.901 | 0.921 |

(d) Sample 4

<table>
<thead>
<tr>
<th>Time</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.877</td>
<td>1.000</td>
<td>0.948</td>
<td>0.991</td>
</tr>
<tr>
<td>Time 3</td>
<td>0.891</td>
<td>0.910</td>
<td>1.000</td>
<td>0.999</td>
</tr>
<tr>
<td>Time 4</td>
<td>0.895</td>
<td>0.915</td>
<td>0.937</td>
<td>1.000</td>
</tr>
</tbody>
</table>

|     | 100.01 | 100.60 | 98.40 | 99.48 |
| M   |       |       |       |       |
| SD  | 22.28 | 26.94  | 31.88 | 35.94 |
| Reliability | 0.857 | 0.9021 | 0.930 | 0.945 |

Note: The numbers in boldface represent the observed between-occasion correlation coefficients, corrected for attenuation using the estimated reliabilities.
here to a repeated-measures analysis of variance (ANOVA). There were no significant main effects for time of observation for any of the four different samples, indicating that there were no significant changes, over time, in introversion. This is not surprising, given that the data were generated so that there would be no change in the average level over time (i.e., the expected values of the slopes are equal to zero).

The observed between-occasion correlation coefficients, corrected for attenuation using the estimated reliabilities, appear in boldface print above the main diagonals in Table 1. Using these corrected correlations as measures of relative stability (see earlier discussion of reliability and estimates of relative stability), one would arrive at the mistaken impression that in each of the four samples, introversion shows almost perfect relative stability.

With four waves of data it is also possible to adopt some of the aforementioned covariance-structure-based strategies to assess relative stability. By assuming that introversion changes according to a first-order autoregressive model (see earlier discussion), it is possible to fit a model to the data using a program for covariance structure analysis (e.g., LISREL). The program estimates the parameters of a prespecified model, and, assuming that the data appear to adequately fit the model, these parameters can be used to assess relative stability.

Specifically, it has been assumed that the true scores on introversion followed a first-order autoregressive model and that the observed scores, therefore, produced a quasi-simplex structure (i.e., a simplex model that allows for measurement errors; Joreskog & Sorbom, 1986). In addition, equal error variances across occasions were assumed.

This model was fit to each of four samples using LISREL (see Joreskog & Sorbom, 1986, pp. 111.70–111.81 for a detailed example). The program successfully produced estimates of the parameters for the autoregressive model for the data from the first three samples. The program was unable to estimate the model for the sample from the fourth population.

Table 2 presents a summary of the goodness-of-fit indexes provided by LISREL for Samples 1 through 3. The autoregressive model appears to fit the data from Populations 2 and 3 fairly well, and, considering the sample size, provides an approximate fit to the data from the first population.

Among the parameters estimated by the program are the standardized regression coefficients (i.e., the $\theta_3$'s) between the latent variables (i.e., the true introversion scores) at each time of measurement. These coefficients are equivalent to test–retest correlations between the corresponding latent variables, and the correlation between any two nonadjacent variables is equivalent to the product of the intervening $\theta_3$'s (Werts, Linn, & Joreskog, 1977). Alternately, these parameter estimates can be viewed as

$^*$This is likely due to the high degree of collinearity among the underlying true scores that was produced by the correlation of 0.9 between the initial score and the slope.
TABLE 2
Summary of Goodness-of-Fit Indexes From LISREL Analyses

<table>
<thead>
<tr>
<th>Sample</th>
<th>GFI</th>
<th>AGFI</th>
<th>RMSR</th>
<th>$x^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.983</td>
<td>0.914</td>
<td>0.011</td>
<td>10.77</td>
<td>0.005</td>
</tr>
<tr>
<td>2</td>
<td>0.992</td>
<td>0.960</td>
<td>0.006</td>
<td>4.88</td>
<td>0.087</td>
</tr>
<tr>
<td>3</td>
<td>0.999</td>
<td>0.993</td>
<td>0.002</td>
<td>0.81</td>
<td>0.668</td>
</tr>
</tbody>
</table>

Note: GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; RMSR = roar-mean-square residual; $t^/$ = chi-square goodness-of-fit test statistics each with 2 dfs; $p$ = probability associated with chi-square test.

correlations between the observed variables, with the attenuation due to the less than perfect reliability of measurement corrected. These parameters, therefore, can be used as estimates of relative stability.

These stability estimates for our simulated introversion data appear in Table 3. They suggest that there is considerable relative stability in introversion over time.

In sum, application of these commonly used approaches to address questions of stability of personality in longitudinal data would lead to the conclusion that there is a high degree of both absolute and relative stability in introversion, each of the hour populations. A somewhat different conclusion is reached when the same data are analyzed from a growth curve perspective.

Growth Curve Approach to the Assessment of Stability

As discussed earlier, growth curve analysis begins with the specification and estimation of a model of change for each respondent. From this per-

TABLE 3
LISREL Estimates of Stability Coefficients

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 2</td>
<td>0.888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.803</td>
<td>0.904</td>
<td></td>
</tr>
<tr>
<td>Time 4</td>
<td>0.757</td>
<td>0.852</td>
<td>0.942</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample 2</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 2</td>
<td>0.912</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.874</td>
<td>0.958</td>
<td></td>
</tr>
<tr>
<td>Time 4</td>
<td>0.860</td>
<td>0.942</td>
<td>0.984</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample 3</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 2</td>
<td>0.931</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.898</td>
<td>0.964</td>
<td></td>
</tr>
<tr>
<td>Time 4</td>
<td>0.887</td>
<td>0.953</td>
<td>0.989</td>
</tr>
</tbody>
</table>
In the present case it was assumed that change in introversion followed a linear model, and, as such, separate linear regression lines were fit to each individual’s observed scores. The means and variances of the parameters from this model (i.e., slopes and intercepts) are reported in Table 4. However, the observed slopes and intercepts are estimated with error. As such, the variances of these parameters are overestimates of the variance of the true parameters (Willett, 1988). One advantage of adopting the growth curve perspective is that when change is assumed to follow a linear model, and when there are more than two waves of data, it is possible, using procedures outlined by Willett (1988, pp. 402-404), to obtain estimates of the error variances, allowing a direct calculation of the variability of the true parameters. Further, once these estimates have been obtained, it is possible to estimate the reliabilities of the individual slopes and intercepts by calculating the ratio of the estimates of the variances of the true values to the corresponding observed values. Estimates of these quantities also appear in Table 4.

In order to assess average levels of absolute stability from the growth curve perspective, one examines the mean value of the individual slopes. As is evident in Table 4, the average value of the individual slopes in each of the four samples is quite low, with none of these average slopes being significantly different from zero. These results accurately reflect the nature

---

Note: Values in parentheses are estimates of variance of true slopes.

---

TABLE 4
Means and Variances for Parameter Estimates From Individual Growth Curves

<table>
<thead>
<tr>
<th>Sample</th>
<th>Statistic</th>
<th>Intercept</th>
<th>Slope*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>97.47</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>415.66</td>
<td>42.38 (29.21)</td>
</tr>
<tr>
<td></td>
<td>Reliability</td>
<td>0.87</td>
<td>0.69</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>98.34</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>436.61</td>
<td>37.80 (22.71)</td>
</tr>
<tr>
<td></td>
<td>Reliability</td>
<td>0.88</td>
<td>0.60</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>99.34</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>389.93</td>
<td>34.91 (21.00)</td>
</tr>
<tr>
<td></td>
<td>Reliability</td>
<td>0.877</td>
<td>0.60</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>99.81</td>
<td>-0.38</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>440.32</td>
<td>38.66 (23.91)</td>
</tr>
<tr>
<td></td>
<td>Reliability</td>
<td>0.90</td>
<td>0.62</td>
</tr>
</tbody>
</table>

The inclusion of error terms not affect the mean values because of the assumption that the expected value of error is equal to zero.
of the populations and are consistent with those obtained from the more traditional analyses reported above (i.e., the repeated measures ANOVAs), in which it was concluded that there was absolute average stability across time.

To assess relative stability from the growth curve perspective, one must first examine the variability of the estimated individual slopes. And, although there is no commonly accepted procedure for testing whether or not a sample variance is significantly different from zero, an examination of the variances of the estimated slopes in Table 4 would seem to suggest a considerable degree of heterogeneity, with this heterogeneity being indicative of between-individual differences in within-individual change (e.g., the rates of change in introversion are different for different individuals).

Given these between-individual differences, relative stability can be assessed by examining the relationship between the estimates of the parameters of the individual growth curves (i.e., the estimates of the slope and the intercept for each individual). The observed correlations between these estimates appear in Table 5. However, these correlations are not only attenuated by the unreliability of the slopes and intercepts, they are, as noted above, negatively biased estimates. In order to obtain unbiased estimates of disattenuated correlation coefficients, a maximum likelihood procedure for estimating the population regression coefficient of the true rate of change (i.e., slope) on the true initial status (i.e., intercept) was applied to the present data (cf. Blomgvisc, 1977). A 95% interval estimate of the true regression coefficient was obtained and the estimated regression coefficients were subsequently converted to correlation coefficients, producing the interval estimates of the correlation coefficient shown in the third column of Table 5. The correlation coefficients between true slopes and true initial value that were calculated directly from the simulated true scores for the various samples used appear in the fourth column of the table.

### Table 5

<table>
<thead>
<tr>
<th>Sample</th>
<th>Observed</th>
<th>Est. Range°</th>
<th>True°</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.035</td>
<td>0.070–0.345</td>
<td>-0.0025</td>
</tr>
<tr>
<td>2</td>
<td>0.259</td>
<td>0.353–0.636</td>
<td>0.313</td>
</tr>
<tr>
<td>3</td>
<td>0.414</td>
<td>0.538–0.816</td>
<td>0.548</td>
</tr>
<tr>
<td>4</td>
<td>0.729</td>
<td>0.900–1.134</td>
<td>0.918</td>
</tr>
</tbody>
</table>

Note: °Estimated range of true correlation coefficient.
*True correlation coefficient obtained from correlation between true scores obtained in simulations.
Inspection of Table 5 reveals that the interval estimates of the true correlations contain the true correlation for Samples 3 and 4, and come close to capturing the correlation in the other two instances. What is perhaps most problematic with this approach is that in all instances the true correlation coefficient is located near, or beneath, the lower limit of the range of estimated true correlations. This suggests that Blomqvist’s (1977) procedure might provide an over correction; however, further work using these procedures would have to be conducted in order to substantiate this view.

Despite this possible shortcoming, it can be argued that the growth curve analyses produced a more accurate description of the nature of the relationships among the underlying true scores than did the more traditional analyses. Specifically, for the samples from Populations I and 2 (in which the correlations between the true slopes and intercepts were set to 0.00 and 0.30, respectively), the interval estimates of the slope–intercept correlations indicated low to moderate relationships between the observed initial level of introversion and the observed races of change, a result consistent with the comparatively low levels of relative stability imposed on the artificial data. For the third sample, the estimated correlation between the slope and intercept ranged from 0.54 to 0.82. This result quite closely reflects the moderate level of stability present in the simulated data (in which the correlation between true slopes and intercepts was set to 0.60). Finally, the interval estimate of 0.90 to over 1.0 for the slope–intercept correlation that was obtained from the growth curve analysis of the sample from the fourth population quite accurately reflected the high degree of relative stability that was present in the artificial data (in which the true initial value—slope correlation was .90). These results stand in strong contrast to the traditional analyses reported earlier, which indicated high levels of relative stability for all four samples.

Summary of Analyses of the Simulated Data

The results presented here suggest that the growth curve analyses provide an accurate description of the nature of the relationships among the true scores. Specifically, these analyses were able to detect high levels of relative stability when such stability was imposed on the underlying distributions from which the samples were obtained. Further, and perhaps more important in terms of assessing stability and change in personality data, the growth curve analyses were comparatively sensitive in terms of detecting heterogeneity in the underlying growth curves when that heterogeneity was in fact present. These results are quite different from the results based on the more traditional approaches that seemed to overemphasise the degree of relative stability that was present in the underlying capacities.

GROWTH CURVE ANALYSES 167
data. It seems warranted to wonder how much of the generally high degree of stability found in personality research is attributable to the method of analysis rather than to true stability.

OTHER BENEFITS OF GROWTH CURVE ANALYSES

The present results suggest the potential usefulness of adopting a growth curve perspective in terms of evaluating stability and change in longitudinal research in general, and in personality data in particular. Further, not all of the strengths of the growth curve approach were emphasized in the data simulations. Among the more common difficulties that arise in collecting real longitudinal data are (a) that some individuals are inevitably unable to be measured on all occasions and (b) that the logistics of collecting longitudinal data frequently preclude the measurement of all individuals at the same time. The first problem has often been dealt with by deleting individuals with missing data (a far from desirable solution), whereas the second problem has often resulted in violations of the time-invariant assumption of certain statistical techniques (e.g., ANOVA). Growth curve analyses can readily handle both of these situations (Laird & Ware, 1982).

Recently, Kenny and Campbell (1989) suggested that the investigator who is interested in assessing questions of change and stability in personality is confronted with the choice between two alternative classes of models: growth curve models and autoregressive models. After some consideration of these two models, Kenny and Campbell concluded that autoregressive models have certain advantages that merit their use. Interestingly, one of the shortcomings noted with regard to the potential use of growth curve analyses is that "growth curve modelers discount individual change and assume perfect stability" (Kenny & Campbell, 1989, p. 448). This conclusion not only seems to be at odds with the present results, but also appears to be a misinterpretation of the essence of growth curve analysis that is fundamentally based on the assessment of individual change.

Further, Kenny and Campbell's (1989) recommendation that investigators adopt autoregressive models is also challenged by the present results. These authors suggested that use of autoregressive models can be predicated, at least in part, on the basis of how well they fit the observed data (Kenny & Campbell, 1989, p. 447). However, unilateral adoption of autoregressive models based solely on their goodness of fit to the observed data seems unwise. As Rogosa and Willett (1985) have noted, and as demonstrated in several of the LISREL analyses in this chapter, data that are generated from nonautoregressive models (as in the constant rate of growth model adopted as the basis for the data simulations here) can often conform to a simplex correlation pattern. Thus, although not wanting to totally discount the usefulness of autoregressive models, we would argue that growth curve
analyses also have their place in the assessment of stability and change in personality—a role that has to date, we believe, been underemphasized.

SUMMARY AND CONCLUSIONS

In this chapter, we have attempted to explain how the techniques of growth curve analysis can be applied to longitudinal personality data. We began by clarifying the definitions of two types of stability, absolute stability and relative stability. The former refers to the amount of stability in the absolute level of a trait over time; the latter refers to the degree of stability in the rank order of the trait over time.

After reviewing several approaches to measuring these two types of stability, we argue that several problems in the measurement of change could best be solved by adopting a different perspective on change than has usually been adopted. Rather than focusing on discrete jumps (or drops) in a measured attribute, we argue that it may be preferable to conceptualize change as a continuous process for each individual. That is, we can think of personality change as forming a continuous function in time for each individual. Growth curve analysis proceeds by attempting to uncover the nature of this function for each individual, and questions about the degree of stability in the population are examined by aggregating the various parameters that describe these individual functions.

Through simulated data, we have shown that several of the methods of analysis discussed were able to accurately capture the level of absolute stability in the data. However, of the methods used, growth curve analysis was the only one that came close to accurately indicating the level of relative stability in the data. These findings suggest that further comparative study of these methods is clearly needed.

In the meantime, however, these findings lead us to suggest that we should be cautious in offering an answer to the question in the title of this book. Although a fairly large research literature supports the notion that personality is stable in adulthood (see Costa & McCrae, this volume), this literature is based largely on data analyzed with the techniques that, if our simulations are to be believed, tend to overestimate the degree of relative stability present.

What conclusion, then, can we reach? We would suggest that caution be exercised when choosing a method of analysis for future research on personality change. When designing studies, it is desirable to obtain measurements on more than two occasions in order to allow the use of growth curve analysis. At the very least, data from studies of personality change

\[1\] More than two measurement occasions are necessary or desirable for many of the other techniques described, as well.
should he analysed with several of these techniques. If different techniques suggest different conclusions about the degree of personality stability, then the question must remain open. However, when multiple techniques yield the same conclusions, then we can truly feel confident about an answer to the question, Can personality change?

REFERENCES


Zimmerman, D. W., & Williams, R. H. (198Z). Gain scores in research can be highly reliable. Journal of Educational Measurement, 19, 149-134.

GRUHT CURVE ANALYSES