Change, Measurement of

Why is the measurement of change over time so important in educational research? The answer is straightforward. When people acquire new skills, when they learn something new, when they grow intellectually and physically, when their attitudes and interests develop, they are changing in fundamental and interesting ways. By being able to measure change over time, it is possible to map phenomena at the heart of the educational enterprise. Education is intended to foster learning, to bring about changes in attitude, achievement, and values. Only by measuring individual change is it possible to document each person’s progress and, consequently, to evaluate the effectiveness of educational systems.

Unfortunately, despite its obvious substantive importance, there has been much controversy over the years about the measurement of change. Influential methodologists concerned about the technical properties of change measurement have argued that it is impossible to measure change well. Empirical researchers were advised to avoid its measurement. Policymakers were told to mistrust research findings based on that measurement.

However, in reality these traditional conclusions and recommendations are incorrect. Change can not only be measured; with a little foresight and planning it can be measured well. Well-publicized false conclusions about the difficulty of change measurement are rooted in a simple misconception: the misconception that individual change should be viewed as an increment, as the difference between “before” and “after.” Early methodologists and empirical researchers saw each person under study as acquiring a quantum of achievement (or attitude, or value, or whatever) between the “pre” and the “post” measurement. They thought that investigators should only be concerned with the size of the acquired chunk. This was a mistaken perception. Individual change takes place continuously over time; it is not purely a “before” and “after” phenomenon. The failure to perceive change as a continuous process of development hamstrung the creation of decent statistical methods for its measurement.

However, methodologists have recently modified their position. They have concluded that individual change can be measured well, providing that research moves beyond the limitations of the “before and after,” or “two wave,” design. They now understand that the continuous process of human development can be easily and efficiently documented if each person is measured repetitively over extended periods of time. In this entry the new “multiwave” perspective is presented, and it is argued that, with a little planning, all questions about change can be answered. Because there has been such controversy, this entry begins by reconsidering traditional methods for the measurement of change.

1. Traditional Methods for Measuring Change

Implicit in any discussion of change are two important kinds of question. The first treats each person as an individual and asks about within-person change. For example, in an evaluation of different reading curricula, it can be asked: Does this person’s reading achievement improve over time? Is there rapid improvement, or is improvement more gradual? Does reading achievement grow at a steady rate, or does it level out after one or two years?—and so on for every person in the sample. The second type of question asks about between-person differences in change. Do certain types of people change in different ways? Does the reading achievement of girls change faster than that of boys? Does progress in reading achievement differ by type of first-grade reading program? These are questions about the way in which individual change is related, over people, to background, training, environment, and so forth.

Answering questions about within-person change logically precedes the asking of questions about between-person differences in change. It is necessary to know how each person changes before asking whether the individual changes differ in some systematic way from one person to the next. However, the latter questions about between-person differences in change often have the greatest practical importance in educational research.

Traditionally, in research on change, investigators collected only two waves of data on each person in the sample. They observed each person’s “status”—their achievement, attitude, or other attribute, say—at the beginning and the end of the investigation. Then, to answer questions about within-person change, they used the “pre” and “post” data to construct a summary measure of individual change for each person in the study. Subsequently, when questions about between-person differences in change were addressed, the two-wave change summaries were simply regressed on, or correlated with, other variables (the “predictors of change”) describing background, training, or treatment.

This seems straightforward, but the apparent simplicity masks serious pitfalls. For one thing, for the strategy to be effective the researcher must be able to use the two waves of data that have been collected to construct a decent measure of change for each person in the sample. Methodologists differ as to how this should be done and psychometric history is littered with strategies proposed for the purpose. Some of these strategies are better than others, but they are all inferior to the multiwave methods introduced in Sect. 2.

The simplest two-wave measure of change is the observed difference score, obtained by subtracting the initial measurement from the final one for each person. Originally the difference score was highly favored but then fell into disrepute, being much
maligned in the 1960s and 1970s. This led to the birth of a coterie of alternative two-wave measures of change, including regression-based estimators of true change and residual change scores. Although the use of two-wave change measurement is not advocated in practice, it is desirable to comment briefly below on these measures in order to clarify the continuing controversy.

1.1 Two-wave Measures of Within-person Change

Because of the uncompromising vicissitudes of nature, when a test or rating instrument is administered the obtained measurement combines a measure of the person's true capability and whatever random error happens to accompany measurement. Of course, research interest focuses squarely on true status—a commodity that, if measurement error is large, may differ considerably from observed status. Similarly, when members of a group of people are changing over time on some important attribute, it is not the fallible observed changes that are of critical interest but the underlying true changes. Measures of observed change simply provide a fallible lens through which the hidden nature of true change may be discerned.

Typically, on the ith occasion of measurement (i.e., at time $t_i$), methodologists have been content to assume that observed status, $Y_{ip}$, is simply the sum of true status, $\eta_{ip}$, and measurement error, $\varepsilon_{ip}$ (the second subscript indicating that the pth person is being referenced). These errors of measurement, $\varepsilon_{ip}$, are usually assumed to be drawn independently from a normal distribution with zero mean and variance $\sigma^2$. When two waves of data have been collected, observed “pretest” and “post-test” status (measured at times $t_1$ and $t_2$) can be represented by their requisite sums: $Y_{1p} = \eta_{1p} + \varepsilon_{1p}$ and $Y_{2p} = \eta_{2p} + \varepsilon_{2p}$. Then, the observed difference score for the pth person, $D_p$, obtained by subtracting $Y_{1p}$ from $Y_{2p}$—is the sum of the underlying true change, $\Delta_p = (\eta_{2p} - \eta_{1p})$, and the difference between the two measurement errors, $(\varepsilon_{2p} - \varepsilon_{1p})$.

Statistically speaking, the observed difference score is a reasonable commodity. It is intuitively appealing and easy to compute. It is an unbiased estimator of the underlying true change. Despite these advantages, the difference score has been resoundingly criticized for both its purported unreliability and its correlation with initial status. Although these deficiencies have been shown recently to be largely imaginary, they are discussed here briefly in order to set the record straight.

Some have said that the difference score is always unreliable; others have argued that it cannot be both reliable and valid simultaneously (Bereiter 1963, Linn and Slinde 1977). In general, neither of these claims is correct. They are misconceptions arising from misinterpretation of the concepts of reliability and validity in the context of change over time (Rogosa et al. 1982, Rogosa and Willett 1983, 1985, Willett 1988). Thus, it has been argued that an instrument can only be regarded as measuring the same construct on both occasions of measurement if the correlation of time-1 and time-2 scores is high. This is fallacious. Even when the instrument is perfectly construct-valid, growth may be heterogeneous across people (i.e., different people may be growing at different rates). Then the individual growth trajectories will naturally intersect as time passes, the rank order of people in the group will fluctuate from occasion to occasion, and the between-wave correlation will be less than unity. It is perfectly possible, in the context of highly heterogeneous growth, for the time-1/time-2 correlation to be large and negative even when the instrument is perfectly construct-valid.

As a consequence of misinterpreting the observed pretest–post-test correlation as an index of construct validity in expressions for the reliability of the difference score, authors have mistakenly focused on hypothetical situations in which there is little variation in true change across people. By choosing to examine situations in which the variance of true change is close to zero, they have ensured that the numerator of the expression for the reliability of the difference score (which is defined as the ratio of the variances of true and observed change) is very small and so the reliability is almost zero. In reality, in perfectly ordinary situations, the reliability of the difference score can be quite respectable (Rogosa and Willett 1983). And when variation in true change from one person to the next is large, the reliability of the difference score can even be greater than the reliabilities of the constituent pretest and post-test scores (Willett 1988).

Anyway, even if the difference score were always unreliable, this would not necessarily be a problem for the measurement of within-person change. Low difference-score reliability does not imply unilaterally that within-person change has been measured imprecisely. Low reliability often occurs in practice because most of the people in the sample are changing at about the same rate (especially over the short term). So even though the 20 points (say) that everyone has changed can be measured very precisely, the changes of different people cannot be distinguished from one another and the difference score appears unreliable. This problem of interpretation does not call the difference-score itself into question, since it is possible to know quite precisely that everyone has changed by 20 points. On the contrary, it simply undermines the credibility of reliability as a worthwhile indicator of measurement quality.

The difference score has also been falsely condemned for at least three other reasons, all of which originate in critics’ misunderstanding of the association between it and pretest status (Linn and Slinde
Change, Measurement of

1977, Plewis 1985). If there is a positive correlation between change and pretest status, for instance, then people with high initial status will tend to have high gains subsequently. If the correlation is negative, then people with low initial scores will tend to change more rapidly than those with high initial scores. This has led some critics to claim that any measure of change that is not independent of initial status must be unfair because it gives “an advantage to persons with certain values of the pretest scores” (Linn and Slinde 1977 p. 125). But such prejudice is unreasonable: why should change and status be unrelated? The intimate connection between change and status is a consequence of growth history. Current status is a product of prior change; current change determines future status. A correlation between change and status is an almost inevitable fact of life.

Second, numerous investigators have empirically estimated the correlation between change and initial status and have worried because their findings often disagreed. Even when investigating similar populations in similar settings with the same measures, some have found the correlation to be positive, some zero, and some strongly negative. But why should they expect a single value for this correlation? When different people are changing in different ways, individual trajectories are likely to crisscross with time and the correlation between change and initial status can fluctuate markedly as different times are selected as the occasion for “initial” measurement (see Rogosa and Willett 1985). Unless some important occasion can be agreed upon substantively to be the initial time, researchers should expect to disagree when they ask: What is the correlation between change and initial status? Because the answer, of course, is, “It depends on the time that you define as ’initial’.”

Third, the difference score has been inappropriately criticized because some people have convinced themselves that its correlation with pretest score is always negative. In reality, this claim is false (for an example of a positive correlation between the difference score and change see Thorndike 1966). Those who condemn the difference score in this way have usually committed one of two mistakes. Either they have spuriously created the negative correlation themselves by “standardizing” the pretest and posttest scores to the same standard deviation before computing the difference score (an ill-advised process that destroys legitimate growth information). Or they are simply being confused by the vagueness of their statistical estimation. In this latter case, they have usually thoughtlessly used the sample correlation of observed initial status and the difference score to estimate the population correlation of true initial status and true change (the real correlation of interest). Unfortunately, because the pretest measurement error appears with a negative sign in the difference score, this estimator is negatively biased. It is often negative even when the underlying true correlation is positive. However, the bias is easily corrected (see Willett 1988) and, anyway, as Rogosa et al. (1982) noted, an unbiased estimate of within-person change—the difference score—should not be rejected because it is a poorly conceived estimator of the association between true change and true initial status is biased. It is the latter that needs fixing, not the former.

So even though the difference score is not the outcast that many critics have claimed, several modifications of it have been proposed in order to estimate true change. $\Delta_d$, better for example, Webster and Bereiter’s (1963) reliability-weighted measure of change, and Lord’s (1956) regression-based estimated true change. These modified estimators improve the measurement of within-person change by trading off unbiasedness for a reduction in mean-squared error. Under very broad assumptions, these modified measures are simply weighted linear combinations of the observed difference score for the particular person and the average observed difference score over the entire group, with weights that depend on the reliability of the difference score. Essentially, the weighting scheme places emphasis on those aspects of the change measurement that are the most “trustworthy”—favoring the difference score when it is reliable but, otherwise, setting each person’s change equal to the average change for everyone when there is little real between-person variation in change to be detected. Even though the modified scores are better estimates of true within-person change, they are usually perfectly correlated with the corresponding difference scores and have (almost) identical reliabilities. Therefore, parallel investigations of between-person differences in change using either of these modified scores usually lead to identical conclusions.

The creation of the residual-change score was motivated by an unnecessary desire to create measures of change that were uncorrelated with pretest score. Residual-change scores are obtained by estimating the residuals that would be obtained in the population regression of true final status on true initial status. They are intended to describe the true change that each person would have experienced, if everyone had “started out equal.” Various methods have been proposed for their computation (see Rogosa et al. 1982). Much energy has been dissipated in the psychometric literature detailing the properties of the many estimators of residual change and considerable argument has been aroused. When discussing residual-change scores, methodologists disagree as to exactly what is being estimated, how well it is being estimated, and how it can be interpreted. In addition to the many technical and practical problems that arise in the empirical application of residual change scores, there also remain unresolved issues of logic and substance that are too extensive to detail.
here (but see Rogosa et al. 1982, Rogosa and Willett 1985, Willett 1988). The researcher is strongly advised to avoid these scores as measures of within-person change.

1.2 Detecting Between-person Differences in Change
Once a two-wave measure of change has been computed for each person, the relationship between change and other “background” variables is often investigated by the common methods of correlation and regression analysis. For instance, to find out if changes in achievement are related to gender, the data-analyst might simply correlate pre/post differences in test score with a dummy variable representing gender. Unfortunately, while straightforward, this rudimentary strategy is flawed. As has been seen, the difference score is a fallible measure of change that contains both true change and measurement error. This latter random noise attenuates the between-person findings, causing the obtained sample correlations to underestimate the true relationship between change and the covariates. This problem can be avoided by correcting between-person analyses for the fallibility of the difference score (for methods of correction, see Willett 1988).

However, the disattenuation process requires that additional information be available to the data-analyst, usually in the form of an external estimate of the reliability of the difference score. Furthermore, because the disattenuation is sensitive to minor fluctuations in the magnitude of the reliability estimate and because the quality of these estimates themselves is often dubious, there exists the very real possibility of glorious imperfection.

If the difference score is abandoned and an estimate of true change is used instead, information in addition to the pre/post measurement is still required for the estimate of true change to be constructed in the first place. In fact, regardless of the measure of change adopted, acceptable between-person analyses can only be conducted if the researcher possesses information in addition to the pair of pretest and post-test scores. This emphasizes the real and fundamental weakness of the two-wave design: there is insufficient information in two waves of data to measure individual change well. To do a satisfactory job, multiwave data are required.

2. Modern Methods of Change Measurement
Taking a “snapshot” of a person’s observed status “before” and “after” is not the best way to reveal the intricacies of their progress. Changes might be occurring smoothly over time with some complex and substantively interesting trajectory. Crude pre/post measurement can never reveal the details of that trajectory. To do a good job of describing individual change over time, a truly longitudinal perspective must be adopted.

Figure 1
Observed growth records of two children whose reading achievement was measured in Grades 1 through 4. The included trend-lines were fitted by OLS linear regression analysis, conducted separately for each child.

2.1 Assembling and Inspecting the Observed Growth Record
People must be followed carefully over time, with multiple “waves” of data collected on their status at sensibly spaced intervals. The investigator must assemble an observed growth record for each person in the dataset. If the attribute of interest is changing steadily and smoothly over a long period of time, perhaps three or four widely spaced measurements on each person will be sufficient to capture the shape and direction of the change. But, if the trajectory of individual change is complex, then many more closely spaced measurements may be required (see Willett 1989).

Once collected, preliminary analyses of the observed growth records are aided by plotting an empirical growth-trajectory for each person—a graph of observed status displayed against time, perhaps with some type of fitted or sketched trend-line included on the graph to summarize broadly the person’s observed growth. Separate inspection of the empirical growth trajectory of each person in the dataset then provides evidence as to whether, and how, each person is changing over time. Fig. 1, for instance, presents the empirical growth trajectories of a pair of children whose reading achievement (scale scores on the Comprehensive Test of Basic Skills) was followed yearly from Grade 1 through Grade 4. In both cases, each child’s observed growth in reading achievement is summarized by a linear trend-line, fitted by ordinary least-squares regression analysis. Comparing trend-line and data suggests that a straight line is a reasonable summary of the
Change, Measurement of

observed growth records and empirical growth-trajectories, a more formal multiwave analysis of change requires that a statistical model be chosen to represent individual change over time. Of course, as has been noted earlier, it is each person’s underlying true change that is of critical interest. Consequently, when a statistical model is selected to represent change, it consists of two parts: (a) a “structural” part that represents the dependence of true status on time; (b) a “stochastic” part that represents the random effects of measurement error.

But what does an individual growth model look like? In drawing the empirical growth trajectories in Figs. 1 and 2, a linear or “straight-line” function of time was chosen as a valid representation of individual change in reading achievement between Grades 1 and 4. In this case the observed reading achievement $Y_{ip}$ of child $p$ at time $t_i$ is represented as follows:

$$Y_{ip} = [\pi_{0p} + \pi_{1p}t_i] + \epsilon_{ip}$$  

where the structural part of the model, describing the dependence of true reading achievement $\eta_{ip}$ on time, has been bracketed to separate it from the measurement error $\epsilon_{ip}$.

The structural part of an individual growth model contains unknown constants called the individual growth parameters. The values of these parameters completely determine the trajectory of true change over time. In Eqn. (1), for instance, there are two individual growth parameters: $\pi_{0p}$ and $\pi_{1p}$. The first parameter, $\pi_{0p}$, is the “intercept” of the straight-line growth model representing person $p$’s true status when $t$ is equal to zero (usually defined as the time at which the study began, and therefore $\pi_{0p}$ represents the true initial status of person $p$). The second parameter, $\pi_{1p}$, is the “slope” of the straight-line growth model representing the rate at which person $p$’s true status is growing over time. If $\pi_{1p}$ is positive, then person $p$’s true status is increasing with time; if it is negative then person $p$’s true status is decreasing with time.

Of course, when analyses of multiwave data are conducted, the straight-line model is not the only available representation of individual change. Many different models—perhaps an infinite number—can be hypothesized. Some are simple, like the straight-line model, while others, like the negative-exponential and quadratic curves, are more complex. It is even possible to join several different models together, creating a piecewise individual growth function.

Perhaps the investigator’s most important task is to select the individual growth model that will be used. Usually this decision is made via extensive preliminary data-analyses in which the observed growth records are systematically and carefully explored. However, the choice of a particular model for a specific purpose can often be informed by a theory of psychological, social, or educational devel-

---

2.2 Choosing a Statistical Model to Represent Within-person Change

Corresponding to the exploratory inspection of observed growth, although Child (b)’s datapoints are a little more widely scattered around the fitted line than are the datapoints for Child (a). Child (a)’s reading achievement is growing at more than double the rate of Child (b), with the latter child having the additional disadvantage of scoring lower initially.

The empirical growth trajectories of everyone under investigation can also be collected together in a single picture. This provides a simple and straightforward way of exploring questions about between-person differences in change. Eyeball comparisons across people may suggest systematic differences in trajectory—children in a phonics-based reading curriculum may tend to change more rapidly than those using a sight-reading curriculum; girls may be growing faster than boys, and so forth. Fig. 2, for instance, adds the fitted growth trajectories of 30 more children to the two already displayed in Fig. 1. The observed data themselves have been omitted to avoid clutter but the trend-lines have been coded to indicate whether the child was judged “ready for reading” in the kindergarten year (solid line = high reading readiness; dashed line = low reading readiness).

Inspection of Fig. 2 suggests that children who were rated highly on reading readiness in kindergarten tended to have both higher reading achievement in Grade 1 and more rapid rates of growth between Grades 1 and 4. Similar plots could be created to display the effects of other interesting predictors.
opment. Theory may suggest, for instance, that individual change in a specific domain is constrained to rise to a ceiling. So, by entertaining a growth model that includes an asymptote, the researcher can test this hypothesis and ultimately investigate whether substantively interesting features of the change are associated with differences in people's background, training, and environment. It may be, for instance, that certain background characteristics predict the ultimate limits on growth (the asymptotes), while others predict the rate at which the ceiling is approached. When substantive theory informs model choice, a richer research arena is revealed.

2.3 A Model for Between-person Differences in Change

Once a particular model has been selected to represent true change in a particular domain, everyone in the sample is assumed to have the same generic functional form for their growth. Different people, however, can have different values of the individual growth parameters. For instance, as in the case of Figs. 1 and 2, when within-person change is linear individuals may differ in intercept and slope. Even more interestingly, the individual growth parameters may differ from person to person in a way that is systematically related to variables describing critical attributes of the people involved (the "predictors of change"). Under the straight-line growth model for reading achievement, for instance, the investigator can ask: Is there a systematic relationship between true rate of change in reading achievement and reading readiness in kindergarten?; between true initial status and reading readiness?; between the individual growth parameters and gender?; home environment?; type of reading program?; and so forth.

Hypothesized relationships between individual growth parameters and predictors of change can be formalized in statistical models for between-person differences in change that are similar to the more familiar regression models (see Rogosa and Willett 1985). Eqn. (2), for instance, presents a pair of between-person models, the true intercepts and true slopes of the straight-line reading achievement growth model in Eqn. (1) as a function of a pair of predictors, \( FEMALE \) (0=boy, \( 1=\text{girl} \)) and \( \text{READYREAD} \) (teacher rating of reading readiness in kindergarten):

\[
\begin{align*}
\pi_{0p} &= \beta_{00} + \beta_{01}FEMALE + \beta_{02}\text{READYREAD} + u_{0p} \\
\pi_{1p} &= \beta_{10} + \beta_{11}FEMALE + \beta_{12}\text{READYREAD} + u_{1p}
\end{align*}
\]  

(2)

where \( u_{0p} \) and \( u_{1p} \) represent residuals (the parts of \( \pi_{0p} \) and \( \pi_{1p} \) that are "unexplained" by \( FEMALE \) and \( \text{READYREAD} \)). In these models, the \( \beta \) coefficients summarize the population relationship between the individual growth parameters and the predictors of change. They can be interpreted in the same way as regular regression coefficients. A nonzero value of \( \beta_{01}, \beta_{02}, \beta_{11}, \) or \( \beta_{12} \) indicates that the corresponding covariate is a predictor of true initial status or true rate of change respectively. For instance, if girls tend to grow more rapidly than boys (i.e., if they have larger values of \( \pi_{1p} \)) then \( \beta_{11} \) will be positive. If children who had higher reading readiness scores in kindergarten have higher initial reading achievement in Grade 1, then \( \beta_{00} \) will be positive, and so on. All of the \( \beta \) coefficients in Eqn. (2) are estimated in the between-person phase of a growth study.

One of the main advantages of the longitudinal perspective is that the investigator is not limited to the simple within- and between-person models presented here. Substantively valid curvilinear functions can be used to model within-person growth, and complex between-person models can be used to relate interindividual differences in several growth parameters to many predictors simultaneously. Systematic interindividual variation in the rate parameter of the negative-exponential growth model, or in the acceleration parameter of the quadratic growth model, can be examined. A new universe for measuring change, far from the worlds of the pre/ post design, is revealed.

3. Doing the Statistical Analyses

In an investigation of change, within-person and between-person statistical models—like Eqns. (1) and (2)—are fitted to data and their parameters estimated and interpreted. Many methods of model-fitting and parameter estimation are available. Some are very simple and can easily be implemented on the popular commercially available statistical computer packages; others are more sophisticated and require dedicated computer software. In this section, in order of increasing complexity, a taxonomy of such data-analytic strategies is briefly reviewed.

3.1 Estimating the Within-person Growth Parameters by Ordinary Least-squares Regression Analysis

Once a sensible within-person growth model has been adopted, it can be fitted separately to each of the individual observed growth records by ordinary least-squares (OLS) regression analysis (in the same way that the fitted trend-lines were created for Figs. 1 and 2). This person-by-person growth modeling leads to individual estimates of the several growth parameters that can then become the dependent variables in subsequent second-round between-person data analyses. In the case of the straight-line growth model in Eqn. (1), OLS-obtained intercepts and slopes estimate the true initial status and rate of change for each of the people in the sample and can be related directly to background predictors in follow-up correlation or regression analyses corresponding to Eqn. (2). This strategy is straight-
forward and easy to implement, and the OLS-estimated intercept and growth rates provide more precise measurement of individual change than was possible with the difference score.

3.2 Improving the Between-person Analyses with Weighted Least-squares Regression Analysis

Due to the idiosyncrasies of measurement, some people usually have empirical growth records whose entries are smoothly ordered and for whom the growth data fall very close to the underlying true growth trajectory. Other less fortunate people have more erratic growth records and their datapoints are scattered, deviating widely from the underlying growth trajectories. These differences in scatter affect the precision (the standard errors) with which the first-round growth parameter estimation is carried out. People with “smooth and systematic” growth records will have highly precise growth parameter estimates with smaller standard errors; people with “erratic and scattered” observed growth records will have less precise growth parameter estimates with larger standard errors.

Between-person analyses of the relationship between the estimated growth parameters and predictors of change can be improved (made asymptotically efficient) if the precision of the first-round growth parameter estimates is taken into account. To achieve this, weighted least-squares (WLS) regression analysis can be used to fit the between-person models using weights that are inversely related to the square of the standard errors of the first-round parameter estimates. Then, more precisely estimated individual growth parameters (those with the smallest standard errors) will play a more important role in determining the outcomes of the second-phase analyses. Willett (1988) presented an expression for weights that can be used in conjunction with Eqs. (1) and (2). Any growth analysis will automatically be improved if the first-round estimation of individual growth parameters is more precise. In practice, this is easily achieved by collecting more waves of data on each person in the sample. In the case of straight-line growth, for instance, the standard errors of fitted growth-rates decrease dramatically as extra waves of data are added. It is for this reason alone that multivariate designs provide superior methods for measuring change. Ultimately, researchers can control the quality of their findings: they can simply “add waves” to their design until some desired level of precision is reached (see Willett 1988).

3.3 Estimating the Within-person and Between-person Growth Models Simultaneously Using Dedicated Computer Software

In this entry it has been argued that high-quality measurement of change is possible via the collection of longitudinal data and the fitting of within-person and between-person models such as those in Eqs. (1) and (2). Such models constitute an algebraic “hierarchy” describing the statistical structure of growth data to be analyzed. Methodology for fitting such hierarchical models has advanced rapidly. In the early 1990s there were several dedicated computer programs available for simultaneously estimating all of the parameters of such models, and for providing appropriate standard errors and goodness-of-fit statistics. Kreft et al. (1990) provide a comprehensive description of four of them and compare their workings. They conclude that, in general, “all four programs tend to converge to the same solution” (p. 100). One of them—called “HLM”—has been widely used in the empirical literature and is well-supported (Bryk and Raudenbush 1992).

3.4 Estimating the Within-person and Between-person Models using Covariance Structure Analysis

There have also been advances in applying the methods of covariance structure analysis to the measurement of change. Covariance structure methods are powerful and general and can be applied easily to many problems in a wide variety of settings. Pioneering work on the application of these methods to the estimation and comparison of mean growth curves was conducted by McArdle and his colleagues (McArdle and Epstein 1987) and other authors have demonstrated how these methods can be used to answer questions about predictors of change (Muthén 1991, Willett and Sayer, in press).

See also: Measurement in Educational Research: Intellectual Development across the Lifespan; Human Development: Research Methodology; Cognitive Development: Individual Differences

References


Change, Measurement of

Washington, DC


Willett J B 1988 Questions and answers in the measurement of change. Review of Research in Education 15: 345–422
