Psychology Science, Volume 49, 2007 (4), p. 375-414

Latent growth curve modeling as an integrative approach to the analysis of change

MANUEL C. VOELKLE[1]

Abstract

Latent Growth Curve Models (LGCM) are discussed as a general data-analytic approach to the analysis of change. Conventional, but popular, methods of analyzing change over time, such as the paired *t*-test, repeated measures ANOVA, or MANOVA, have a tradition, which is quite different from the more recently developed latent growth curve models. While the former originated from the idea of variance decomposition, the latter have a factor analytic background. Accordingly, "traditional methods", which focus on mean changes, and "new methods", with their emphasis on individual trajectories, are often treated as two entirely different ways of analyzing change. In this article, an integrative perspective is presented by demonstrating that the two approaches are essentially identical. More precisely, it will be shown that the paired *t*-test, repeated measures ANOVA, and MANOVA are all special cases of the more general latent growth curve approach. Model differences reflect the underlying assumptions, and differences in results are a function of the degree to which the assumptions are appropriate for a given set of data. Theoretical and practical implications are set forth, and advantages of recognizing latent growth curve models as a general data-analytic system for repeated measures designs are discussed.

Key words: Latent growth curve models, paired *t*-test, repeated measures ANOVA, MANOVA, latent difference scores

[1] Manuel C. Voelkle, Chair of Psychology II, Schloss, 68131 Mannheim, Germany; +49 621-181-2131 (voice), +49 621-181-2129 (fax), email: voelkle@rumms.uni-mannheim.de

Almost forty years ago, Cohen (1968) showed that the analysis of variance (ANOVA) and multiple regression analysis are essentially identical data analytic systems. His publication received so much attention among social scientists like few other articles since that time. This was even more surprising, given that the actual message was not new, and the underlying mathematical principles were well known among statisticians. As a matter of fact, it was less the "discovery" itself, but more the theoretical and practical implications that came along with it, which caught the attention of many researchers. A few years earlier, Cronbach (1957), in his presidential address at the Sixty-Fifth Annual Convention of the American Psychological Association, called for an integration of the "two disciplines of scientific psychology" (p. 671): Experimental and Correlational Psychology. Even though the distinction between the two disciplines alludes to more than the use of different statistical procedures, the focus on individual differences made regression techniques particularly interesting to correlational psychologists. Experimental researchers, on the other hand, were typically more interested in group differences, thus preferring the analysis of variance. By demonstrating that ANOVA and multiple regression (MR) yield the same results if group membership is coded as a set of dummy variables in MR, Cohen (1968) provided the methodological basis for an integration of the two disciplines. Today, this is common knowledge among social researchers, even though some introductory statistics texts still treat multiple regression and ANOVA as if these were two completely unrelated techniques. Although not new in statistical content, Cohen's work (Cohen et al., 2003; Cohen & Cohen, 1983; Cohen, 1968) had a tremendous impact on the statistical thinking of many researchers. On the one hand, it showed experimental researchers the limits of the analysis of variance and exemplified the strict assumptions, which are associated with these models, on the other hand it demonstrated the flexibility of multiple regression, while at the same time pointing correlational researchers to the dangers of this flexibility by comparing it to traditional ANOVA techniques. Finally, however, Cohen's work helped integrating two different ways of statistical thinking. The analysis of group differences and individual differences were no longer viewed as fundamentally different research approaches in need of different statistical procedures, but were shown to be closely related. As a result, researchers not only gained a better understanding of the strengths and weaknesses of their preferred statistical approach, but psychological research in general moved towards a fusion of its two disciplines (Cronbach, 1975; Cook & Campbell, 1979).

Today, we find a similar situation in the analysis of change. On the one hand, there are the "traditional approaches" dealing with the analysis of mean changes, on the other hand there are the "new methods for the analysis of change" (Collins & Sayer, 2001) focusing on individual changes over time. Both classes comprise an entire family of different models, with the repeated measures ANOVA and Latent Growth Curve Models (LGCM) being the two most prominent representatives of either class.

The ANOVA for repeated measures was developed as a direct extension of the fixed-effects techniques of the analysis of variance pioneered by Fisher in the 1920s and '30s (e.g., Fisher, 1925). As shown in Equation (1), the basic idea is to partition the total sum of squares (SS_{Total}) into one part that is caused by interindividual differences ($SS_{Between}$) and one part that is due to intraindividual changes over time (SS_{Within}). Below we will come back to Equation (1), for now it suffices to recall that this allows us to control for systematic but often unwanted between-subject variance.

$$SS_{Total} = SS_{Within} + SS_{Between} \tag{1}$$

What is not apparent from Equation (1) is the fact that this approach depends on a number of strict assumptions. Beside the usual assumptions for the analysis of variance, such as multivariate normality, independence, and homogeneity of covariance matrices, it is primarily the assumption of sphericity, which is often not met in practice. As a consequence, alternative procedures, such as the multivariate analysis of variance (MANOVA), have been proposed for analyzing repeated measures. Although less restrictive, MANOVA is a direct extension of the analysis of variance and rests upon the same underlying idea of variance decomposition. Central to both approaches is their focus on group changes instead of individual changes. The separation of between and within-subject variance is merely a means to the end of controlling for differences between subjects in order to partition the remaining within-subject variance into variation due to potential covariates (SS_A) and variance not accounted for (SS_{Error}). As shown in Equation (2), only group mean differences $(\overline{x}_t - \overline{x}_\bullet)$ for time point $t = 1 \ldots T$ are of interest, while all person ($i = 1 \ldots n$) specific deviations are treated as error variance.

$$SS_{Within} = SS_A + SS_{Error}$$
$$SS_A = n\sum_{t=1}^{T}(\overline{x}_t - \overline{x}_\bullet)^2 \quad \text{and} \quad SS_{Error} = \sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ti} - \overline{x}_i - \overline{x}_t + \overline{x}_\bullet)^2 \tag{2}$$

As will be discussed below, this approach is often not only overly restrictive, but also ignores valuable information contained in the data.

While interindividual differences in intraindividual change are treated as error variance in traditional methods, they are of primary interest in latent growth curve modeling. LGC-models have their roots in factor analysis, going back to Tucker (1958) and Rao (1958) who proposed the application of factor analytic techniques to the organization of individual growth curves (see McArdle & Nesselroade, 2002, or Bollen & Curran, 2006, for a brief history of LGCM). Attempts to model individual growth curves can be found even earlier (Wishart, 1938), a major breakthrough, however, was the publication of Meredith and Tisak (1990; 1984). Using a slightly different notation than the present paper, they demonstrated that individual change over time can be expressed as a structural equation measurement model (Equation (3)), while interindividual differences in intraindividual change correspond to the latent variable structural model (Equation (4)). In standard structural equation modeling (SEM), the measurement model relates the observed variables to the latent factors by assuming that the former can be expressed as an exclusive function of the latter. The structural model on the other hand, allows formulating and testing explicit hypothesis regarding the relationship among the latent variables. This distinction constitutes the basis of the common LISREL (LInear Structural RELations) notation, which is also adopted in the present paper. An excellent introduction is provided by Bollen (1989, p. 10-39). In case of repeated measures, the T points of measurement are represented by the $T \times 1$ vector $\mathbf{x}$. Accordingly, $\boldsymbol{\tau}$ is a $T \times 1$ vector of intercepts and $\boldsymbol{\varepsilon}$ a $T \times 1$ vector of person and time point specific error terms. $\boldsymbol{\eta}$ is an $m \times 1$ vector of (growth) factors with the $T \times m$ factor loadings matrix $\boldsymbol{\Lambda}$. As illustrated in Equation (4), the latent factor(s) can be regressed on other exogenous or en-

dogenous variables (represented by the $h \times 1$ vector $\boldsymbol{\xi}$, respectively the $m \times 1$ vector $\boldsymbol{\eta}$) weighted by the $m \times h$ matrix $\boldsymbol{\Gamma}$, respectively the $m \times m$ matrix $\mathbf{B}$. Analogous to Equation (3), $\boldsymbol{\alpha}$ is an $m \times 1$ vector of intercepts and $\boldsymbol{\zeta}$ an $m \times 1$ vector containing the error terms. Equations (3) and (4) will be discussed in more detail further below. The resulting approach to the analysis of change is very general, and as noted by Meredith and Tisak (1990) "with imagination and careful attention to detail, given suitable identification, every form of repeated measures ANOVA or MANOVA can be built up as a special case" (p. 114).

$$\boldsymbol{x} = \boldsymbol{\tau} + \boldsymbol{\Lambda\eta} + \boldsymbol{\varepsilon} \tag{3}$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\Gamma\xi} + \boldsymbol{B\eta} + \boldsymbol{\zeta} \tag{4}$$

By demonstrating how to use common methods of covariance structure analysis to analyze individual growth curves, they prepared the ground for present-day latent growth curve models. Even though the technique has been extended during the last decade, the mathematical basis is still the same. With some exaggeration, one could even say that there are little advancements that were not envisioned in the original Meredith and Tisak (1990) paper. This also applies to the present article, where no large claim of originality is being made. As a matter of fact, most of the material presented herein has already been published in some scattered articles or chapters. However, I am not aware of any systematic discussion of the conditions and consequences of integrating traditional analysis of variance techniques into a general LGC-modeling framework. Typically, "traditional" methods to analyze change and latent growth curve models are discussed separately, thereby emphasizing their differences instead of their commonalities. In our view, however, much can be learned about either approach by taking a closer look at their interrelationship. Latent growth curve modeling must not be viewed as just another "tool in the toolbox of methods", but should be understood as a very general data analytic system for repeated measures designs which incorporates paired *t*-tests, repeated measures ANOVA, and MANOVA as special cases. We hope that this article will help to evoke a similar "new look" (Cohen & Cohen, 1983, preface) on the analysis of change as Cohen's (1968) seminal article on multiple regression/correlation analysis forty years ago.

Overview of the article

The article has three sections and a concluding discussion. In the first section I begin with the analysis of two-wave data and demonstrate how the paired samples *t*-test can be viewed as a special case of a latent growth curve model. Emphasis will be put on conceptual differences between change scores, residualized (true) gain scores and latent difference scores. In section two, I extend the discussion to multi-wave data by contrasting repeated measures ANOVA, MANOVA and LGCM. The underlying assumptions of each approach will be highlighted and advantages of LGCM to analyze change will be discussed. Section three deals with different ways to predict change and provides a comparison across methods. I conclude with a discussion of the theoretical and practical implications of latent growth curve modeling as a general data analytic system.

Two-wave data

Two repeated points of measurement are the minimum requirement for the analysis of change. Although two time points do not constitute a real *longitudinal* study (Rogosa, Brandt, & Zimowski, 1982; Singer & Willett, 2003), the simple pre-post-test is probably one of the most often used research designs in experimental research. For example, one might be interested in the effectiveness of an intervention, or improvement on a learning task, where the performance of each individual has been assessed at the beginning and at the end. Table 1 shows the scores of $n = 17$ female and $n = 18$ male participants on a hypothetical learning task, where performance has been assessed on four consecutive equidistant time points (x_1 to x_4). The data will be used to illustrate the main arguments throughout the remainder of this article. Each score might correspond to the average number of points obtained and points lost in a computer based complex problem-solving scenario. Typical examples of such tasks are TRACON or ATC (e.g., Ackerman, 1992; Ackerman & Kanfer, 1993). However, because the data are artificial and were chosen only for illustrative purposes, the reader is welcome to think of any other (learning) task. Ignoring any possible differences between men and women at the moment, one of the most basic questions is whether the average performance is significantly better at the end of the task than at the beginning. This question can be easily addressed by a paired samples *t*-test. For this purpose, one would compute the mean $\bar{d} = 1/N \sum (x_4 - x_1)$ of the difference d between x_1 and x_4. Under the assumption that d is roughly normally distributed, the ratio of $\bar{d}$ to its standard error constitutes the well-known *t*-test for repeated measures as shown in Equation (5).

$$t = \frac{\bar{d} - 0}{\frac{SD_d}{\sqrt{N}}} \tag{5}$$

For $\bar{d}$ = 4.660 - 1.112 = 3.549 and estimated standard deviation $sd_d = 1.423$, the test statistic $t = 14.750$ is highly significant in this example ($df = 34$, $p < .01$)[2]. Computing the difference between pre- and post-test corresponds to a separation of between- and within-person variance as shown in Equation (1). By subtracting initial performance from final performance, interindividual differences ($SS_{Between}$) are kept constant and the analysis concentrates on the within-subject variation (SS_{Within}). As illustrated in Table 2, the paired *t*-test is identical to a one factor repeated measures ANOVA, which will be discussed later on.

[2] To minimize the problem of rounding errors, we will report all results with a precision of up to three decimal places. Most computations, however, will be made with a higher precision. This may result in some minor inconsistencies in the text, but will prevent us from carrying along rounding errors and will improve overall precision.

Table 1:
Example data set of a hypothetical learning task with four repeated points of measurement (x_1 - x_4) and two predictors (g and sex)

Subject	x_1	x_2	x_3	x_4	g	sex
1	2.28	2.28	2.81	4.49	97.32	M
2	0.12	2.26	2.94	5.42	99.95	M
3	0.46	1.47	1.63	2.95	88.71	F
4	0.52	2.03	2.76	3.69	94.31	F
5	0.01	1.72	3.08	3.56	93.76	F
6	0.11	1.98	3.71	6.94	119.01	M
7	1.10	2.19	2.21	5.00	95.31	F
8	0.77	2.03	2.91	4.33	99.51	F
9	1.06	1.64	1.94	1.97	84.48	F
10	2.85	2.57	3.13	4.85	94.19	M
11	1.28	2.54	2.59	3.18	96.29	F
12	2.36	3.58	6.28	8.15	95.76	M
13	2.15	3.84	4.26	6.34	111.21	M
14	1.04	2.57	3.15	3.00	104.6	F
15	1.70	3.52	5.24	5.85	99.82	M
16	1.14	1.53	2.55	2.45	109.44	F
17	0.45	2.12	1.71	4.41	99.83	F
18	0.75	2.28	3.83	4.20	94.98	M
19	0.44	2.79	5.35	5.77	105.06	M
20	1.68	3.12	4.21	5.51	87.19	M
21	1.93	3.22	4.08	5.40	110.82	M
22	1.15	1.75	1.83	1.36	87.71	F
23	1.85	2.82	3.97	5.09	101.72	M
24	0.38	0.85	2.86	3.14	98.57	F
25	1.90	2.56	3.28	3.75	108.68	M
26	0.91	1.91	3.00	5.67	87.62	M
27	1.72	2.75	3.70	5.24	100.65	F
28	0.82	2.88	4.20	5.09	102.93	M
29	0.37	2.00	3.38	5.00	103.51	F
30	0.82	2.44	2.18	5.22	97.84	F
31	0.73	1.84	2.45	4.38	102.03	F
32	1.54	2.91	4.38	5.07	105.19	M
33	1.03	2.97	4.83	6.07	110.35	M
34	0.60	2.69	3.54	4.84	119.71	M
35	0.89	2.51	3.81	5.73	100.3	F
Mean	**1.11**	**2.41**	**3.37**	**4.66**	**100.24**	
Sd	**0.71**	**0.65**	**1.08**	**1.40**	**8.40**	

Total: covariances (correlations)

	x_1	x_2	x_3	x_4
x_1	0.500 *(1.00)*			
x_2	0.274 *(.601)*	0.418 *(1.00)*		
x_3	0.259 *(.339)*	0.525 *(.754)*	1.162 *(1.00)*	
x_4	0.222 *(.224)*	0.604 *(.666)*	1.111 *(.734)*	1.971 *(1.00)*

Women: covariances (correlations)

	x_1	x_2	x_3	x_4
x_1	0.178 *(1.00)*			
x_2	0.106 *(.517)*	0.238 *(1.00)*		
x_3	0.027 *(.096)*	0.122 *(.377)*	0.444 *(1.00)*	
x_4	-0.024 *(-.046)*	0.331 *(.543)*	0.410 *(.492)*	1.566 *(1.00)*
Mean	0.817	1.999	2.632	3.801

Men: covariances (correlations)

	x_1	x_2	x_3	x_4
x_1	0.664 *(1.00)*			
x_2	0.216 *(.492)*	0.291 *(1.00)*		
x_3	0.072 *(.095)*	0.357 *(.713)*	0.861 *(1.00)*	
x_4	-0.025 *(-.030)*	0.218 *(.398)*	0.610 *(.648)*	1.030 *(1.00)*
Mean	1.390	2.788	4.058	5.472

Note: Sd = Standard Deviation; F = Female (F = 0), M = Male (M = 1).

Table 2:
Paired samples *t*-test, one factor repeated measures ANOVA and LGC-model for two time points (x_4 *and* x_1) and no predictors.

		Paired samples *t*-test	
Mean difference (x_4 - x_1)	*t*	*df*	*p*
3.549	14.750	34	.000
		Repeated measures ANOVA	
Source	*F*	*df*	*p*
Time	217.568	1	.000
		LGCM	
	Estimate (α_1)	*Standard error (SE)*	*p*
	3.549	0.237	.000

Note: $t^2 = F$ for $df_{\text{numerator}} = 1$.

A latent growth curve approach to the analysis of two-wave data

The *t*-test can also be specified as a structural equation model as graphically illustrated by Figure 1A. By fixing all factor loadings, we essentially realize the assumption of classical test theory (CTT) that an observed score is the sum of a true score and an error component (Gulliksen, 1950; Lord & Novick, 1968). In Equation (6) this assumption is illustrated for the first point of measurement (x_1), where η_0 refers to the true score and ε_1 to the error at time point one.

$$x_1 = \eta_0 + \varepsilon_1 \tag{6}$$

Applying the same assumption to x_4 (i.e., $x_4 = \eta_4 + \varepsilon_4$ with $\eta_4 = \eta_0 + \eta_1$) and solving for η_1, Equation (7) is obtained by simple algebraic transformations.

$$\eta_1 = (x_4 - \varepsilon_4) - (x_1 - \varepsilon_1) \tag{7}$$

Obviously, η_1, as specified in this model, maps true intraindividual change from pre- to post-test (Steyer, Eid, & Schwenkmezger, 1997). Looking at the *t*-test from this perspective points to another crucial assumption of the conventional *t*-test, that is the absence of any unsystematic (measurement) error. As a matter of fact, setting the variances, covariances and means of all error terms to zero – as indicated by the dotted lines in Figure 1A – is necessary for the SE-model to be identified. Using the general matrix notation introduced in Equation (3) and (4), the *t*-test can be expressed as a special case of a latent growth curve (SEM) model with

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_4 \end{pmatrix} \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_0 \\ \eta_1 \end{pmatrix} \quad \boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \quad \boldsymbol{\tau} = \boldsymbol{\varepsilon} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

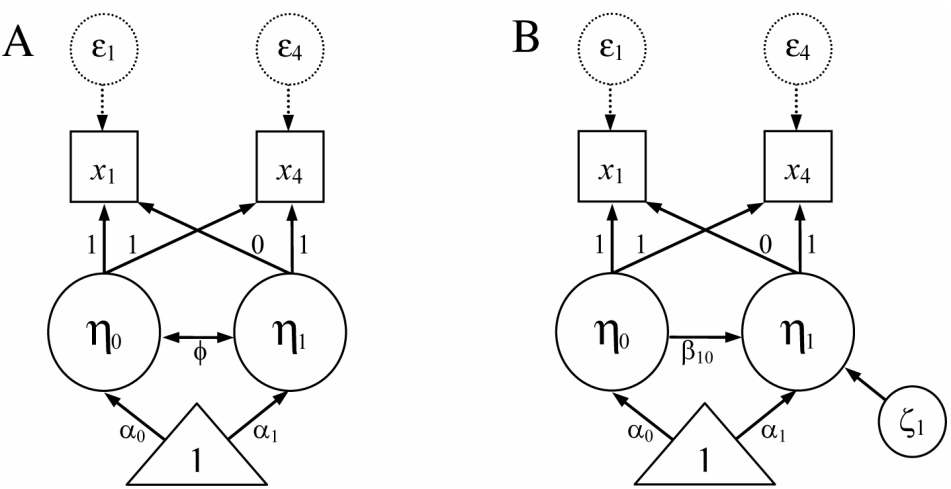

Figure 1:
Path diagram of a paired samples *t*-test (A) and a base-free measure of change model (B). The triangle represents the constant 1. Accordingly, the two regression weights α_0 and α_1 are the means of the two latent factors η_0 and η_1. ϕ represents their covariance. The dotted error terms (ε_1 and ε_2) indicate that the model does not account for measurement error (mean(ε_1) = mean(ε_2) = $\sigma(\varepsilon_1) = \sigma(\varepsilon_2) = 0$)

Since no predictors of change are considered in this section, $\boldsymbol{\xi}$, $\boldsymbol{\Gamma}$, $\mathbf{B}$ and $\boldsymbol{\zeta}$ simply drop out of Equation (4). Allowing the covariance (ϕ) between η_0, and η_1 to be freely estimated, the resulting model is just identified ($df = 0$) and the critical ratio of $\alpha_1 = 3.549$ to its standard error ($SE = 0.237$) is asymptotically identical to the *t*-value of the paired samples *t*-test reported above[3,4].

Change scores, residualized gain scores and latent difference scores

In case the reliability (r_{tt}) of the measurement instrument(s) would be known, adopting the SEM approach allows us to take this knowledge into account by fixing the variance of the $t = 1 \ldots T$ error terms to $var(\varepsilon_t) = (1 - r_{tt}(x_t)) * var(x_t)$. As long as $E(\varepsilon_t) = 0$, a comparison of means via the paired *t*-test or LGCM, would still yield identical results, while all higher moments, such as the variances and covariances of the two latent variables, will differ. Especially when analyzing predictors and correlates of change, this has some profound implications and as pointed out by Raykov (1999), modeling change on a latent dimension is often a

[3] The structural equation modeling software Mplus (Muthén & Muthén, 1998-2007) was used for the estimation of all LGC-models throughout the paper.

[4] Maximum likelihood (ML) estimation was used, thus the critical ratios follow approximately a *z*-distribution and results will be asymptotically identical.

better approach than modeling observed change scores. To elaborate on this point, consider Equation (8), which defines the reliability of change scores $\rho_{tt}(d)$ as a function of the reliability of the pretest $\rho_{tt}(x)$ and the reliability of the post-test $\rho_{tt}(y)$.

$$\rho_{tt}(d) = \frac{\sigma_x^2 \rho_{tt}(x) + \sigma_y^2 \rho_{tt}(y) - 2\sigma_x \sigma_y \rho_{xy}}{\sigma_x^2 + \sigma_y^2 - 2\sigma_x \sigma_y \rho_{xy}} \tag{8}$$

ρ_{xy} denotes the correlation coefficient between pre- and post-test and σ the standard deviation. Based on this formula, the simple difference score has been vehemently criticized and researchers have even been advised to avoid the gain score altogether and "frame their questions in other ways" (Cronbach & Furby, 1970, p. 80). The reason for this lies in the fact that in order to obtain reliable difference scores, the reliability of the pre-test *and* the reliability of the post-test must be high, while at the same time their correlation should be low. If one of these conditions is not met, the reliability $\rho_{tt}(d)$ will be low, so that "the difference score between two fallible measures is frequently much more fallible than either" (Lord, 1963, p. 32). Especially the last condition of a low pre-post-test correlation has caused some confusion about the meaning of gain scores, known as the "reliability-validity paradox". As Bereiter (1963) pointed out, a low correlation between pre- and post-test indicates that different constructs are being measured and as soon as we cannot be sure that we are measuring the same thing, there is no point in analyzing change over time. As a consequence, a number of different strategies have been proposed to somehow correct or improve the gain scores prior to investigating any correlates or predictors of change (see Cronbach & Furby, 1970). The most popular approaches are probably the residualized observed difference score (DuBois, 1957) and the base-free measurement of change (Tucker, Damarin, & Messick, 1966), which I will come back to below (see also Raykov, 1992, 1993a, 1993b).

Eventually, however, it was a series of papers by Rogosa et al. which heralded a reorientation in the analysis of change (Rogosa et al., 1982; Rogosa & Willett, 1983, 1985; Rogosa, 1988). By demonstrating that "many of the deficiencies that have been attributed to differences scores in the behavioral sciences literature are a result of misunderstandings" (p. 730), Rogosa et al. (1982) took on the defense of the difference score. Their arguments are well documented and shall not be repeated at this point (but see Rogosa et al., 1982, pp. 730; Rogosa, 1995). Based on their arguments, it is now clear that the general criticism on the difference score was completely unwarranted (Willett, 1997, p. 215). Rogosa et al. (1982, p. 728) carefully distinguished between true change and observed change and refocused the analysis of change on the individual by employing a linear growth model as shown in Equation (9), which is essentially a simple case of Equation (3).

$$\omega_i(t) = \eta_{0i} + \eta_{1i} t \tag{9}$$

$\omega_i(t)$ is the *true* score of person *i* at time point *t*. For just two measurements (x_1 and x_4), η_1 is identical to the difference between the two *true* scores ω_4 and ω_1 as demonstrated in Equation (7). As a matter of fact, if the reliability of both measures is known and accounted for, the variance of the latent slope factor η_1 is identical to the variance of gain scores corrected for attenuation. Based on Lord (1956, Formula 8), McNemar (1958, Formula 3) demon-

strated that the variance of the true difference scores can be defined as shown in Equation (10).

$$\sigma^2_{\omega_{Diff}} = (\sigma^2_1 + \sigma^2_4 - 2\rho_{14}\sigma_1\sigma_4) - (\sigma^2_{\varepsilon_1} + \sigma^2_{\varepsilon_4}) \tag{10}$$

$\sigma^2_{\omega_{Diff}}$ denotes the variance of the *true* difference (ω_{Diff}) between x_1 and x_4, ρ_{14} is their correlation, and $\sigma^2_{\varepsilon_t}$ the error variance at time point t. In the example introduced above, d was the difference between x_4 and x_1 with $\bar{d} = 3.549$ and $sd_d = 1.423$. Given the observed correlation $r_{14} = 0.224$ (see Table 1) and assuming a reliability of $r_{tt}(x_1) = .80$ and $r_{tt}(x_4) = .85$, we obtain an estimate of $\sigma^2_{\omega_{Diff}}$ = (0.500 + 1.971 - 2 * 0.224 * 0.707 * 1.404) - (((1 - .80) * 0.500) + ((1 - .85) * 1.971)) = 1.630. This is equivalent to the variance of η_1 in the general latent growth curve model with error terms fixed to (1 - $r_{tt}(x_t)$) * $var(x_t)$ as discussed above[5].

As pointed out by Tucker et al. (1966), the variance of the true difference scores can be further partitioned into variance of true independent (base-free) change scores and true dependent change score variance. The latter depend entirely on the pre-test, while the former are entirely independent of it. Even though Rogosa et al. (1982) have warned researchers to exercise "extreme caution" (p. 741) when using and interpreting residual change measures, it may sometimes be important to distinguish between change which would have occurred if everyone started out equal, and change which is a direct function of the pre-test[6]. The variance of true independent gain scores σ^2_γ can be computed in two steps. First, the observed post-test scores are regressed on the pre-test scores, divided by the reliability of the pre-test, to obtain the unstandardized regression coefficient a (see Tucker et al., 1966, p. 462 & technical appendix).

$$a = \frac{\rho_{14}\sigma_4}{\rho_{tt}(x_1)\,\sigma_1} \tag{11}$$

Second, given a, σ^2_γ can be computed as shown in Equation (12).

$$\sigma^2_\gamma = \rho_{tt}(x_4)\sigma^2_{x4} - 2a\rho_{14}\sigma_{x1}\sigma_{x4} + a^2\rho_{tt}(x_1)\sigma^2_{x1} \tag{12}$$

---

[5] As before, the equivalence holds only asymptotically because maximum likelihood estimation was used for the LGCM estimation. In the present case, $sd(\eta_1) = 1.254$ (LGCM-ML) and the estimated $\sigma_{\omega_{Diff}} = 1.277$ (Equation 10). The asymptotic equivalence can be better evaluated by using the covariance matrix provided in Table 1 – instead of the raw data – as input for Mplus. This allows the user to specify any number of observations, without changing the actual information contained in the data. Using a sufficiently large sample size (e.g., 3500) the results differ by less than three digits after the decimal point.

[6] As will be discussed in the next section, the residualized gain scores (or more generally speaking the covariance between intercept and slope) are a direct function of the position of the intercept in time (Rovine & Molenaar, 1998; Stoel & van den Wittenboer, 2003; Biesanz et al., 2004). Especially when the time point of the pre-test is arbitrary (as it is often the case in multi-wave studies), this must be taken into consideration when interpreting residualized gain scores.

In our example, a = (0.224 * 1.404) / (0.80 * 0.707) = 0.556, so that an estimate of σ_γ^2 = 0.85 * 1.971 - 2 * 0.556 * 0.224 * 0.707 * 1.404 + 0.556^2 * 0.80 * 0.500 = 1.552. Asymptotically, the same base-free measure of change can be obtained by regressing the latent (true) difference factor η_1 on η_0. This can be easily done by extending the structural equation model specified above by setting

$$\boldsymbol{B} = \begin{pmatrix} 0 & 0 \\ \beta_{10} & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\zeta} = \begin{pmatrix} 0 \\ \zeta_1 \end{pmatrix}$$

The unbiased variance of the disturbance term $var(\zeta_1)$ is equal to the variance of the base-free measure of change (σ_γ^2) as proposed by Tucker et al. (1966). Figure 1B shows a path diagram of the model.

To summarize, it has been shown that the paired samples *t*-test is a special case of the general latent growth curve model. If the reliabilities of the pre- and post-test are known, LGCM allows the computation of latent difference scores, equivalent to gain scores corrected for attenuation as proposed by Lord (1956) and McNemar (1958). In addition, it is possible to distinguish between true dependent and true independent (base-free) gain scores as originally suggested by Tucker et al. (1966). Usually, however, reliabilities are not simply known, but must be estimated and a good theory is imperative for doing so. Traditionally, reliability estimates were obtained based on the principles of classical test theory (Gulliksen, 1950; Lord & Novick, 1968) by using retests, parallel tests, or various estimates of internal consistency. The often inadequate adoption of CTT to the analysis of change was probably one of the main reasons for difference scores to fall into disgrace in the early seventies (Cronbach & Furby, 1970). Clearly, a measurement instrument which exhibits high retest reliability cannot be suitable for assessing change over time, and it is problematic to define reliability of change indirectly via a lack of stability as done in the traditional Formula (8) (see also Wittmann, 1997, 1988). Naturally, this also applies to the analysis of change via latent growth curve models. However, other than the use of observed difference scores, LGCM provides the flexibility to specify a model of change, which best fits the underlying theory of change. Basically, there are two ways to incorporate theory into our model in order to obtain true change scores (Raykov, 1999). Either multiple indicators must be employed at each time point and theory dictates the specification of the construct in question, or more than two time points must be observed and theory dictates the nature of change over time. In the first case, at least two indicators are required for model identification, in the second case at least three time points must be available. For a more detailed discussion of the former approach see Raykov (1992). Especially the latter approach, however, opens up a variety of different models of change, which will be discussed in the next section.

Multi-wave data

A latent growth curve approach to repeated measures ANOVA

Having at least three time points, the repeated measures ANOVA is probably one of the most often employed statistical procedures for the analysis of change. It is implemented in all major statistical packages and its basic idea is comparatively easy to understand. Especially applied researchers, however, are often unaware of the strict (and oftentimes unrealistic) assumptions the repeated measures ANOVA rests on. Conceiving repeated measures ANOVA as a special case of a more general latent growth curve model not only helps to gain a better understanding of the assumptions underlying ANOVA, but also points to (new) ways how to test and cope with violations of standard assumptions.

As introduced in Formula (1), the repeated measures ANOVA decomposes the total variance additively into variation due to interindividual differences between subjects ($SS_{Between}$) and individual differences within the same subject (SS_{Within}). On a conceptual level it is important to realize that repeated measures ANOVA assumes a *single* variable with a single total variance (i.e., SS_{Total}) which is decomposed instead of *multiple* variables, which is the idea underlying MANOVA. The univariate conceptualization of change implies that a potential covariance between average interindividual differences and interindividual differences in intraindividual change are not part of the model. Analogous to the paired *t*-test (see the comparison of simple gain scores versus residualized gain scores), this is not to say that such a covariance may not exist, it is just not part of the analysis, because of the assumption of a single variable. However, it is precisely this covariance which often not only exists, but is the cause for several logical, statistical and conceptual confusions (Lohman, 1999).

As shown in Equation (2), the SS_{Within} can be further partitioned into variation due to systematic change over time (SS_A) and remaining error variance (SS_{Error})[7]. Returning to our example data set of Table 1 with four time points (x_1 to x_4), a repeated measures ANOVA yields $SS_{Total} = SS_{Within} + SS_{Between} = 288.858 + 85.356 = 374.214$. Table 3 shows the results of the full analysis. The same analysis can be carried out as a special version of a latent growth curve model. For this purpose, the (measurement) error variance-covariance matrix $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}$ is again assumed to be zero, and the original variables are transformed by a contrast matrix $\boldsymbol{\Lambda}$. The last three columns of $\boldsymbol{\Lambda}$ correspond to (T - 1) polynomial contrasts. For the simple case of equally spaced time points, orthogonal polynomials are usually the default in most statistics programs (such as SPSS). The weights (i.e., factor loadings) of each orthogonal polynomial sum to zero and they are mutually independent. The specific weights of polynomial contrasts depend on the number of time points. Note, however, that polynomial contrasts are just a special case of more general contrast codes, which could be imposed via $\boldsymbol{\Lambda}$. Choosing the correct contrast matrix would also permit the analysis of unequally spaced points of measurement. An excellent introduction to power polynomials in general and orthogonal polynomials in specific is provided by Cohen et al. (Cohen et al., 2003, pp. 196). As apparent from the $\boldsymbol{\tau}$-vector, the intercepts of the original variables are freely estimated, while the means of the transformed (latent) variables are all constrained to zero ($\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}} = \boldsymbol{0}$; $\boldsymbol{\alpha} = \boldsymbol{0}$) . For the

[7] Other, and maybe more useful, decompositions are possible but shall not be discussed in this paper (but see Cattell, 1966; Wittmann, 1988).

present example with four repeated measures, the model is defined as shown below and as graphically depicted in Figure 2.

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_0 \\ \eta_1 \\ \eta_2 \\ \eta_4 \end{pmatrix} \quad \boldsymbol{\Lambda} = \begin{pmatrix} 0.5 & -0.671 & 0.5 & -0.224 \\ 0.5 & -0.224 & -0.5 & 0.671 \\ 0.5 & 0.224 & -0.5 & -0.671 \\ 0.5 & 0.671 & 0.5 & 0.224 \end{pmatrix} \quad \boldsymbol{\tau} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{pmatrix} \quad \boldsymbol{\Phi} = \begin{pmatrix} \sigma^2_{\eta_0} & & & \\ \phi_{10} & \sigma^2_{\eta_1} & & \\ \phi_{20} & \phi_{21} & \sigma^2_{\eta_2} & \\ \phi_{30} & \phi_{31} & \phi_{32} & \sigma^2_{\eta_3} \end{pmatrix}$$

Table 3:
Repeated measures ANOVA

Source	***SS***	***df***	***MS***	***F***	***p***	***G-G***	***H-F***
Within(Time)	236.514	3	78.838	153.625	.000	.000	.000
Within(Error)	52.345	102	0.513				
Within Subjects	288.858	105	2.751				
Between Subjects	85.356	34	2.510				
Total	374.214	139	2.692				
Polynomial Contrasts (Within)							
Linear	235.735	1	235.735	215.300	.000		
Error(linear)	37.227	34	1.095				
Quadratic	0.000	1	0.000	0.000	.990		
Error(quadratic)	8.916	34	0.262				
Cubic	0.778	1	0.778	4.267	.047		
Error(cubic)	6.201	34	0.182				

Note: *SS* = Sum of Squares; *MS* = Mean Squares; *df* = degrees of freedom; *G-G* = Greenhouse-Geisser; *H-F* = Huynh-Feldt.

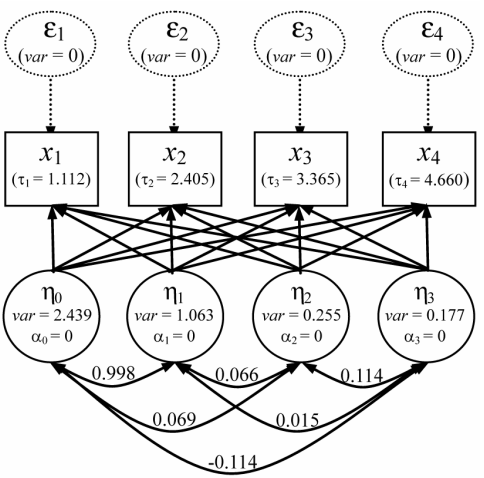

Figure 2:
Path diagram and parameter estimates of a repeated measures ANOVA/MANOVA

For the moment, let us ignore the unconstrained matrix $\mathbf{\Phi}$ with variance σ_η^2 and covariance ϕ between the latent variables, but I will come back to this important point further below. The transformation (factor-loading) matrix $\mathbf{\Lambda}$ is chosen in a way that places the first factor η_0 at the "center" of the observed time period (Wainer, 2000). The unbiased variance of η_0 corresponds to the interindividual differences at the time point where the factor loadings of all other factors are zero. In standard latent growth curve modeling this is usually the first point of measurement, but it can be easily changed as illustrated by the present example (see also Rovine & Molenaar, 1998; Stoel & van den Wittenboer, 2003; Biesanz et al., 2004). After centering the first factor, it maps interindividual differences in average performance, thus it is equivalent to the between-subject variance of the repeated measures ANOVA. As a matter of fact, multiplying the variance of η_0 = 2.439 by 35 (the number of participants) we obtain the $SS_{Between}$ = 85.36 reported in Table 3. Because the intercepts of the manifest variables match their observed means, the group mean differences due to time (SS_A) can be computed as shown in Equation (2). Unfortunately, standard SEM software does not automatically provide this computation, but it can be easily done by hand. The estimated variance of the four means is 2.253, so that $SS_A = n * (T - 1) * 2.253 = 35 * 3 * 2.253 = 236.514$, which corresponds to the SS_A contained in Table 3. Analogous to 35 times the variance of the first factor, which corresponds to the $SS_{Between}$, the sum of the (35 times) the variance of the remaining three latent variables corresponds asymptotically to the SS_{Error} (35 * 1.063 + 35 * 0.255 + 35 * 0.177 = 52.325) as shown in Table 3 and Figure 2. The sum of squares within subjects is now readily computed by adding SS_A and SS_{Error} (SS_{Within} = 236.514 + 52.325 = 288.84). Another way to compute the sum of squares within is to constrain the intercepts of the observed variables to zero ($\boldsymbol{\tau} = 0$, see Figure 3). Now, the sum of (35 times) the variance of the three last latent variables corresponds no longer to the SS_{Error}, but the SS_{Within} (35 * 7.792 + 35 * 0.255 + 35 * 0.200 = 288.65). Finally, the total variance is obtained by adding $SS_{Between}$ and SS_{Within} (85.36 + 288.84 = 374.20).

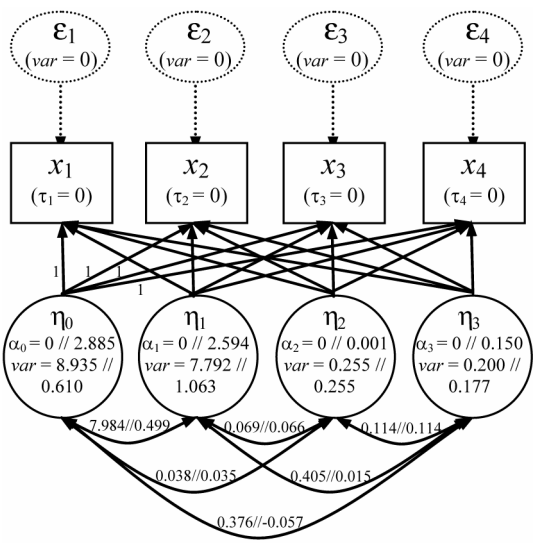

Figure 3:
The same ANOVA/MANOVA model as shown in Figure 2 with all means of the latent variables constrained to zero versus (//) freely estimated. Intercepts ($\boldsymbol{\tau}$) of all manifest variables are fixed to zero.

Trend analysis

Knowing that there are significant changes in means across time is often just a first step towards a more detailed analysis of this change. Thus, the analysis of variance is usually complemented by a trend analysis using single degree of freedom polynomial contrasts (e.g., Cohen et al., 2003, p. 219 & 575). The goal is thereby to examine which function (i.e., linear, quadratic, cubic, etc.) provides the best description of the changes in means. For this purpose, the person (i) and time point (t) specific scores (x_{it}) are regressed on the predictor time ($t = 1 \ldots T$). To obtain orthogonal contrasts, t is transformed into t^* so that the condition $\sum_{t=1}^{T}(t_t^* - \bar{t}^*) = 0$ is met, with $\bar{t}^*$ being the mean of t^*. The transformation is done via the matrix $\mathbf{\Lambda}$ as introduced above. A more detailed introduction to trend analysis is provided by Maxwell and Delaney (2000, pp. 207). The lower part of Table 3 contains the T-1 polynomial contrasts for the repeated measures analysis using any conventional statistical software package. Figure 3 shows the according LGCM path diagram. The factor loadings of each of the three last latent variables (η_1-η_3) correspond to the three orthogonally transformed predictors "time" (λ_t), with $\lambda_0 = 1$ as shown in Equation (13). Setting $\lambda_0 = 1$ does not change the interpretation of the polynomial contrasts, but only the variance and covariances of η_0 as apparent when comparing Figure 2 and Figure 3. This choice was made in order to stay consistent with the output of most statistics programs, where α_0 is treated as a normal intercept (weighted by one). Note, however, that now the variance of η_0 no longer corresponds directly to the between-subject variance.

$$x_{it} = 1^* \alpha_0 + \alpha_1 * \lambda_1 + \alpha_2 * \lambda_2 + \alpha_3 * \lambda_3 + \varepsilon_{it} \tag{13}$$

As for all LGC-models, this requires a reconsideration of researchers familiar with traditional confirmatory factor analysis, since factor loadings are not regression weights but correspond to the predictors, weighted by the means (i.e., fixed regression coefficients) of the latent variables. Readers only familiar with hierarchical linear modeling HLM (e.g., Bryk & Raudenbush, 1992) may find this notion far less confusing. In the same way the means of the latent factors in Figure 3 correspond to the regression weights of a polynomial function, the according sum of squares can be computed by comparing the variance of the latent factors in a model where all means have been constrained to zero to a model where all means have been freely estimated. As discussed above and illustrated by Figure 3, the sum of squares within (SS_{Within}) which can be explained by a linear mean trajectory (SS_{Linear}) is 35 * var_{η_1} (*constrained*) – 35 * var_{η_1} (*unconstrained*) = 35 * 7.792 - 35 * 1.063 = 235.515 = SS_{Linear}, where *constrained* refers to the model with $\boldsymbol{\alpha} = \mathbf{0}$ and *unconstrained* to the model where all means are freely estimated. The same is true for the quadratic ($SS_{Quadratic}$ = 35 * 0.255 - 35 * 0.255 = 0.000) and cubic (SS_{Cubic} = 35 * 0.200 - 35 * 0.177 = 0.805) polynomial contrasts. Naturally, the sum of squares of the three orthogonal polynomial factors add up to the total sum of squares within subjects explained by time ($SS_{Within(time)}$ = 235.515 + 0.000 + 0.805 = 236.32, which corresponds to the $SS_{Within(time)}$ = 236.514 reported in Table 3. As before, minor differences between the LGCM results and the repeated measures ANOVA are in part due to the different estimation procedures and in part due to rounding errors. Knowing all sum of squares and the according degrees of freedom, F-tests can be computed as shown in Table 3. Given the usual assumptions (primarily normal distribution of the observed variables), this

test is asymptotically equivalent to the squared critical ratio (c.r.) of the means provided by standard SEM software. The critical ratio is computed by dividing the parameter estimate (in this case the mean) by its standard error. In our example c.r.(α_1) = 2.594 / 0.174 = 14.887, c.r.(α_2) = 0.013, and c.r.(α_3) = 2.112. Asymptotically, the critical ratios follow a *z*-distribution, so we find that $p(\alpha_1) < .01$, $p(\alpha_2) > .05$ and $p(\alpha_3) < .05$. Apparently, a straight line describes the changes in means very well but there also appears to be a slightly cubic trend. Squaring the critical ratios, we get close to the *F*-ratios obtained by computing standard polynomial contrasts as shown in Table 3.

An alternative (new) approach to significance testing of the polynomial contrasts would be to compare the unconstrained model as shown in Figure 3 to a restricted model where the mean of a single latent variable has been constrained to zero (e.g., $\alpha_1 = 0$). If the observed data are normally distributed, (*n*-1) times the maximum likelihood fitting function (e.g., Bollen, 1989, pp. 107) approximates a χ^2 distribution with degrees of freedom equal to the degrees of freedom of the model in question. The difference of two χ^2 values follows again a χ^2 distribution with degrees of freedom equal to the difference of the degrees of freedom of the two models. Because the unconstrained model shown in Figure 2 and Figure 3 is just identified, it fits the data perfectly, thus χ^2(unconstrained) = 0 and *df*(unconstrained) = 0. In order to test for a linear mean trajectory, α_1 would have to be constrained to zero, resulting in a χ^2(constrained) of 69.729 with *df*(constrained) = 1. The difference χ^2(constrained) - χ^2(unconstrained) = 69.729 - 0 = 69.729, with *df* = 1 - 0 = 1, is highly significant ($p < .01$). The same significance tests can be conducted for a quadratic and cubic trajectory (χ^2(quadratic) = 0.000, $p > .05$, and χ^2(cubic) = 4.199, $p < .05$). Despite the fact that the results are very similar in this example, it must be emphasized – once again – that the likelihood-ratio (i.e., χ^2-difference) approach is a large-sample method as compared to the finite-sample method of comparing the sum of squares (Raykov, 2001). Although the likelihood-ratio approach may offer some advantages over traditional tests, it is unclear whether (and under which conditions) it is appropriate for small samples. Future research is needed to address this issue.

Overall significance tests and underlying assumptions

In the same way significance tests are conducted for the *T*-1 polynomial contrasts, the changes in means over time can be tested for significance. In standard repeated measures ANOVA this is readily done by computing $F = (SS_{Within(time)} / (T - 1)) / (SS_{Within(Error)} / (T - 1) * (n - 1))$ as shown in Table 3. This (univariate) test is identical to a comparison between the abridged (i.e., $(T - 1) \times (T - 1)$) covariance matrix $\mathbf{\Phi}$, with all means being constrained to zero and the unconstrained matrix. Equation (14) shows the computation of the univariate *F*-test, with $\mathbf{\Phi_R}$ denoting the mean-constrained covariance matrix and $\mathbf{\Phi_F}$ denoting the unconstrained (free) matrix as shown in Figure 3 (separated by // in Figure 3). The trace (tr()) of a matrix is the sum of all elements in the main diagonal (i.e., the sum of squares within).

$$F = \frac{(tr(\boldsymbol{\Phi_R}) - tr(\boldsymbol{\Phi_F}))/(T-1)}{tr(\boldsymbol{\Phi_F})/(N-1)*(T-1)} \tag{14}$$

In our example, $tr(\mathbf{\Phi_R}) = 8.246$ and $tr(\mathbf{\Phi_F}) = 1.494$, so that $F = 153.63$, which corresponds to the univariate F-ratio provided in Table 3, which was obtained using any major statistical software package.

A more detailed discussion of Equation (14) will be provided below. At this point, however, it is important to have a closer look at the within-subject variance-covariance matrix $\mathbf{\Phi}$, which I have deliberately ignored so far. In order to be a meaningful statistic (i.e., to follow an F-distribution), the computation of F depends on the assumption of homogeneity of treatment difference variances, which is identical to the assumption of sphericity. Sphericity implies the equality of the variances of the differences between all pairs of repeated measures. If this assumption is not met, the mean differences over time (i.e., $SS_{Within(time)}$) would have to be qualified based on the changes in variance and thus would not be a reasonable estimate of the overall time effect. As a result, the p-values would be biased, leading to an inflated type I error. In practice, the assumption of sphericity is often equated with the assumption of compound symmetry. Compound symmetry exists if the variance-covariance matrix of the repeated measures contains the same elements on its main diagonal (equal variances) and the same elements off the main diagonal (equal covariances). If all variances are equal and all covariances are equal (possibly different from the variances), the variances of the differences between all possible pairs of repeated measures must be equal too. As a matter of fact, compound symmetry is a special case of sphericity, that is if the assumption of compound symmetry is met, the assumption of sphericity is also met and the F-ratio follows an exact F-distribution. However, there are cases where the observed measures do not exhibit compound symmetry, but the sphericity assumption is still met, and the repeated measures ANOVA F-test remains correct[8].

As demonstrated by Raykov (2001), both assumptions can be tested via structural equation modeling and as will be shown below, even the – usually more complicated – test of sphericity is quite easily conducted within the general LGCM framework. Let $\mathbf{\Sigma}$ be the $T \times T$ covariance matrix of the repeated measures and let ρ denote a correlation coefficient, then $\mathbf{\Sigma}$ should equal

$$\sigma^2 \begin{pmatrix} 1 & & \\ \rho & \ddots & \\ \rho & \rho & 1_T \end{pmatrix}$$

if the assumption (H_0) of compound symmetry is met. The alternative hypothesis (H_1) is readily formulated by removing the restriction of equal variances and covariances. In order to test whether the assumption of compound symmetry holds in the present example (Table 1), we would maintain the model as shown in Figure 2, but set

[8] Huynh and Feldt (1970) speak of Type S and Type H matrices and provide an example of a matrix meeting the assumption of sphericity but not the assumption of compound symmetry.

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and compare} \quad \boldsymbol{\Phi} = \begin{pmatrix} \sigma^2 & & & \\ \phi & \sigma^2 & & \\ \phi & \phi & \sigma^2 & \\ \phi & \phi & \phi & \sigma^2 \end{pmatrix}$$

against $\boldsymbol{\Phi}$ (i.e., $\boldsymbol{\Sigma}$) as defined above. In other words, it is tested whether the assumption holds that the repeated measures have the same variance and equal covariances across time. This results in a χ^2(-difference) of 66.014 with 8 degrees of freedom, which is highly significant ($p < .01$), suggesting that the assumption of compound symmetry is not met. As a consequence, the *F*-test of a standard repeated measures ANOVA would not be correct. However, because compound symmetry is only a sufficient but not necessary assumption of the repeated measures ANOVA *F*-test, researchers are better advised to test directly for violations of sphericity, even though some authors argue that this distinction is hardly ever relevant in applied research (Maxwell & Delaney, 2000, p. 473). The same way we can test for deviations from compound symmetry, we can test for deviations from sphericity (see Raykov, 2001). The only difference is that this test refers to the orthogonally transformed variables, instead of the untransformed variables. As introduced above, the orthogonal transformation (actually orthonormal transformation with respect to η_1 - η_3) is implemented by the choice of $\boldsymbol{\Lambda}$. Sphericity exists if the variance-covariance matrix of the *T*-1 transformed variables contains no off-diagonal elements and only equal variances on the main diagonal (i.e., H_0: $\boldsymbol{\Phi} = \sigma^2 * \mathbf{I}$, with $\mathbf{I}$ being a $(T - 1) \times (T - 1)$ identity matrix). In the present case, we would maintain $\boldsymbol{\Lambda}$ as shown below,

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0.5 & -0.671 & 0.5 & -0.224 \\ 0.5 & -0.224 & -0.5 & 0.671 \\ 0.5 & 0.224 & -0.5 & -0.671 \\ 0.5 & 0.671 & 0.5 & 0.224 \end{pmatrix}, \text{ and compare } \boldsymbol{\Phi} = \begin{pmatrix} \sigma^2_{BS} & & & \\ \phi_{BS1} & \sigma^2 & & \\ \phi_{BS2} & 0 & \sigma^2 & \\ \phi_{BS3} & 0 & 0 & \sigma^2 \end{pmatrix}$$

against $\boldsymbol{\Phi}$ with all elements being freely estimated (see Figure 3). Note that the $(T - 1)$ orthonormally transformed variables must meet the assumption of sphericity, while a different variance (σ^2_{BS}) and covariance (ϕ_{BS}) is permitted for the between-subject factor. In the present example this results in a χ^2(-difference) of 45.664 with 5 degrees of freedom, which is again highly significant ($p < .01$), suggesting that the assumption of sphericity is not met. Having worked out the transformation matrix ($\boldsymbol{\Lambda}$), which is provided by most statistic programs or can be looked up in any standard statistics text book, the above test is as easily implemented as the test of compound symmetry. Therefore I see no reason why one should settle for second best (i.e., testing the assumption of compound symmetry), but recommend testing directly for sphericity. As mentioned above, this test is a large sample test, and its performance is not very well known in finite samples such as the present one. Especially for large samples, however, the test may constitute an interesting alternative to Mauchly's criterion *W* (Mauchly, 1940; see also Mendoza, 1980), which tests the assumption of independence and homoscedasticity of the transformed repeated measures. Mauchly's criterion *W* is defined as shown in Equation (15), with $\boldsymbol{\Sigma}$ being the sample covariance matrix of the un-

transformed variables with $df = n - 1$, and $T - 1$ being again the number of orthogonal contrasts.

$$W = \frac{\left|\Lambda' \Sigma \Lambda\right|}{\left(\frac{tr(\Lambda' \Sigma \Lambda)}{(T-1)}\right)^{(T-1)}} \tag{15}$$

For $f = ((T^2 - T) / 2) - 1$ and $d = 1 - (2 * T^2 - 3 * T + 3) / (6 * (T - 1) * (n - 1))$, the product $-(n - 1) * d * \ln(W)$ follows approximately a central χ^2 distribution with f degrees of freedom if $\mathbf{\Phi}$ meets the assumption of sphericity (e.g., see Huynh & Feldt, 1970, p. 1588). In the present case, $W = 0.271$ and the according $\chi^2 = 42.706$ with $f = 5$ degrees of freedom for

$$\mathbf{\Lambda} = \begin{pmatrix} -0.671 & 0.5 & -0.224 \\ -0.224 & -0.5 & 0.671 \\ 0.224 & -0.5 & -0.671 \\ 0.671 & 0.5 & 0.224 \end{pmatrix}$$

and $\mathbf{\Sigma}$ as shown in Table 1 (covariance matrix). Again the assumption of sphericity must be rejected ($p < .01$). Although the results are fairly similar, future research is necessary to provide a better comparison of the traditional Mauchly's test (Mauchly, 1940) and the LGCM likelihood ratio-approach introduced above.

Regardless of which test is being used, it is obvious that the data do not meet the assumption of sphericity and the F-test must not be trusted. As a matter of fact, the repeated measures ANOVA F-test is quite sensitive against violations of the sphericity assumption (e.g., Vasey & Thayer, 1987; Keselman & Rogan, 1980) and it is important to take appropriate action (e.g., see the three step approach of Greenhouse & Geisser, 1959; Keselman et al., 1980). For this purpose, a number of adjusted univariate tests have been developed. The three most prominent approaches are probably the Geisser-Greenhouse lower bound correction, Box's $\hat{\varepsilon}$ adjustment, and the Huynh-Feldt $\tilde{\varepsilon}$ adjustment. All three of them are based on a correction of the degrees of freedom for the critical F value. A more detailed description is beyond the scope of this article, a good overview, however, is provided by Maxwell and Delaney (2000, pp. 475).

A latent growth curve approach to multivariate analysis of variance

Even though the adjustments are a simple and effective way to deal with violations of the sphericity assumption, the principle problem of how to interpret any effects in the presence of variance and covariance changes over time remains. The multivariate approach to the analysis of variance (MANOVA) offers a solution to this problem. As mentioned above, MANOVA assumes several different variables (instead of a single variable whose variance is decomposed in within- and between-subject variance), which may very well exhibit different correlations among each other. As a matter of fact, all models introduced so far (see Figure 2

and Figure 3) are actually MANOVA models because all elements in $\mathbf{\Phi}$ were freely estimated. Other than the test statistics computed in repeated measures ANOVA, the multivariate approach explicitly accounts for these variances. Returning to Equation (14), which shows the computation of the univariate F-test, we find that by taking the trace of $\mathbf{\Phi}$, all off-diagonal elements have been ignored in the ANOVA approach (i.e., all covariances were assumed to be irrelevant). The multivariate analog of Equation (14) simply replaces the trace by the determinant, thus taking into account all elements of $\mathbf{\Phi}$. Equation (16) shows the multivariate test statistic based on the same constrained and unconstrained matrices as described above (e.g., Maxwell & Delaney, 2000, p. 589).

$$F = \frac{(|\boldsymbol{\Phi}_R| - |\boldsymbol{\Phi}_F|)/(T-1)}{|\boldsymbol{\Phi}_F|/(N-T+1)} \tag{16}$$

In our present example (see Table 1), $|\mathbf{\Phi_R}| = 0.259$ and $|\mathbf{\Phi_F}| = 0.034$, so that $F = 71.560$ ($df_{numerator} = 3$, $df_{denominator} = 32$), indicating that there are indeed significant mean changes over time ($p < .01$). By considering all variances and covariances among the (transformed) measures, the F-ratio no longer depends on the assumption of sphericity[9]. This is a major advantage, because the assumption of sphericity is not only very restrictive, but often unrealistic and hardly ever met in the behavioral sciences. Other than the corrections, which are only approximate, the multivariate approach offers an exact test of differences in means over time. Accordingly, the type I error rates are correct even if the assumption of sphericity is violated. On the downside, the traditional repeated measures approach has greater power to detect any potential effects if the assumption of sphericity is met. A more comprehensive comparison of the two approaches to the analysis of change is provided by Maxwell and Delaney (2000, chapter 13).

The LGCM likelihood-ratio test for polynomial contrasts proposed above can be easily generalized to a test of *any* changes in means over time and constitutes a (new) alternative to the multivariate F-test of Equation (16). The according test statistic corresponds to the χ^2-fit of a LGC-model as shown in Figure 3, where the means of all three growth factors (η_1 - η_3) have been constrained to zero ($\alpha_1 = \alpha_2 = \alpha_3 = 0$). In our example $\chi^2 = 71.483$ with $df = 3$, which is again highly significant ($p < .01$), indicating that there are significant mean changes over time. As before, the likelihood-ratio test may be an interesting alternative in large samples, given its ease of implementation, and its independence of the sphericity assumption.

While the relative advantages and disadvantages regarding type I and type II error rates of the repeated measures ANOVA and MANOVA are comparatively well known, future research is necessary to evaluate the performance of the likelihood-ratio test. Especially for many time points the new approach appears to be promising, since it allows the specification of any within-subject covariance structure. For example, it would be possible to implement the assumption of what we might term "partial sphericity", that is the assumption of sphericity for a part of $\mathbf{\Phi}$ but not the entire matrix as required by the multivariate approach. In other words, the LGCM approach offers all advantages of MANOVA regarding potential viola-

[9] While the multivariate approach does not depend on sphericity, it assumes multivariate normality which is – strictly speaking – more restrictive than the assumption of univariate normality underlying the standard repeated measures ANOVA. However, for practical purposes – and other than the assumption of sphericity – this difference is typically negligible.

tions of sphericity, while being more flexible by offering the option to impose specific constraints on the within-subject matrix. This should result in a decrease of type II errors, while type I error rates and the correctness of parameter estimates should remain unchanged.

The latent growth curve modeling perspective

Putting the within-subject covariance matrix into the center of interest is probably the biggest difference – and at the same time the greatest advancement – of LGCM over traditional techniques. In repeated measures ANOVA the focus lies only on mean changes, while the remaining within-subject variance is viewed as error variance and is assumed to have a very restrictive form. The MANOVA approach is more flexible with respect to the nature of the within-subject covariance matrix ($\mathbf{\Phi}$), but the matrix is still treated as pure error (co)variance. In LGCM, however, this matrix is of central interest, because it maps individual changes over time as well as interindividual differences in individual changes. In other words, the focus is shifted away from mean changes towards changes of individual units (i.e., persons). To illustrate this point, consider a factor loading matrix $\mathbf{\Lambda}$ as defined below.

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{pmatrix}$$

This factor loading matrix corresponds to the standard LGCM setup, where η_0 maps true interindividual differences at the first point of measurement (i.e., the latent intercept is positioned at the first point of measurement). The mean of the second latent variable (η_1) corresponds to the average linear increase from one time point to the next, the mean of η_2 maps the average quadratic increase and η_3 the average cubic increase. The changes in means can be described as shown in Equation (13)), with estimated $\alpha_0 = 1.112$, $\alpha_1 = 1.681$, $\alpha_3 = -0.500$, and $\alpha_4 = 0.111$. For instance, the mean of the last point of measurement $t = 4$, would be predicted to be $\tau_4 = 1.112 + 3 * 1.681 + 9 * -0.500 + 27 * 0.111 = 4.66$. Because the model is saturated, the predicted mean is identical to the sample mean as shown in Table 1. In contrast to Equation (13)) where any interindividual differences are contained in the error term (ε_{it}), the interindividual differences in intraindividual change over time are now mapped by the variance of the three growth factors. Figure 4A illustrates this variation. While we see a significant increase in mean performance over time, the increase is also characterized by large interindividual differences. As demonstrated above, this within-subject variance is simply treated as error variance in traditional approaches. However, from Figure 4A it should also be apparent that the within-subject variance is unlikely to be unsystematic as assumed by the repeated measures ANOVA. As a matter of fact, the fan-spread pattern of increasing variance observed in Figure 4A is quite typical for learning data in the behavioral sciences (Kenny, 1974; Campbell & Erlebacher, 1970). While the fan-spread effect does not present a problem for the multivariate approach to the analysis of changes, MANOVA still focuses on changes in means instead of providing a closer investigation of the within-subject covariance matrix.

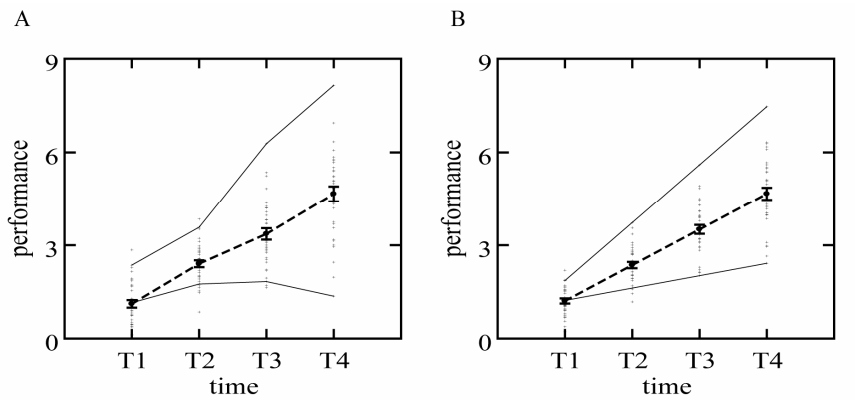

Figure 4:
Performance on a hypothetical learning task as introduced in Table 1. The points along the dotted line indicate mean performance, as well as the corresponding standard error, at each time point. The two solid lines represent the trajectories of the best and worst individual at the last point of measurement. (A) Saturated (descriptive) model (B) estimated linear model

This, however, is readily accomplished by latent growth curve modeling. As described above, the LGC approach can be used to test the assumption of compound symmetry or sphericity (for additional tests see Raykov, 2001), but it is also possible to test much more refined hypotheses. For example, one might be interested in testing the degree to which the data follow a specific trajectory over time and to what extent individuals deviate from the average trajectory. In other words, LGCM offers great flexibility in testing very specific hypotheses regarding change. While this can result in quite complex models, the most basic latent growth curve models are actually very parsimonious, requiring much fewer parameters to be estimated than standard MANOVA models. Especially applied researchers are often not aware of this fact. Other than traditional techniques, however, LGCM buys its advantages from the existence of a good theory. If no prior theory of mean changes and/or individual changes can be formulated, LGCM might indeed have little value over traditional methods. If, however, some prior theory exists, rival hypotheses can be formulated and explicitly tested against each other. As an example, we might suspect that the learning "curves" of the 35 individuals contained in Table 1 and depicted in Figure 4A can be sufficiently well described by a straight line. A standard latent growth curve model as shown in Figure 5, and as defined below, constitutes such a test.

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad \boldsymbol{\tau} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_0 \\ \eta_1 \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \quad \boldsymbol{A} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \quad \boldsymbol{\Phi} = \begin{pmatrix} \sigma^2_{\eta_0} & \\ \phi_{10} & \sigma^2_{\eta_1} \end{pmatrix} \quad \boldsymbol{\Theta}_{\varepsilon} = \begin{pmatrix} \sigma^2_{\varepsilon_1} & \\ 0 & \sigma^2_{\varepsilon_2} \end{pmatrix}$$

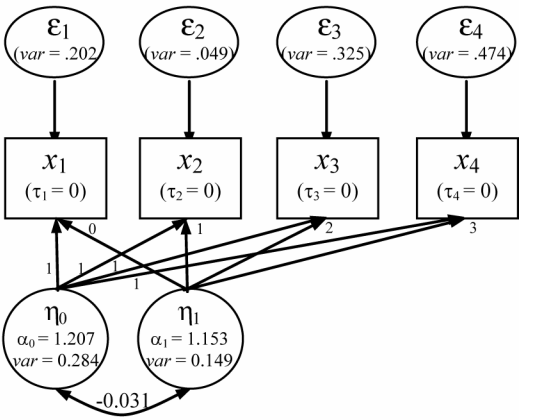

Figure 5:
Linear latent growth curve model

The most obvious difference to all previous models is that the variances of the error terms are no longer constrained to zero but are freely estimated. By imposing a theory of change on the data, the within-subject variance-covariance matrix $\mathbf{\Phi}$ could be further partitioned into mean changes over time, systematic individual deviations from the average (linear) trajectory, and time point specific residual variances represented by $\mathbf{\Theta_\varepsilon}$. It is important to note that the partitioning of the within-subject variance into "systematic" variance contained in $\mathbf{\Phi}$ and "error" variance ($\mathbf{\Theta_\varepsilon}$) is contingent on the underlying theory of change. In that sense it is difficult to distinguish between systematic time point specific variance and pure measurement error, or more generally to distinguish between reliability and validity of change. Figure 4B shows the *estimated* individual trajectories using a linear LGC-model. If the usual assumptions of structural equation models are met (primarily multivariate normality) and sample size is large enough, the resulting model fit provides a test of the goodness of approximation of the estimated trajectories to the observed trajectories shown in Figure 4A. The evaluation and interpretation of fit indices works as usual and will not be reviewed in this paper (but see Bollen & Long, 1993; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Note, that it is also possible to compare competing models of change (nested and non-nested, see Levy & Hancock, 2007). For example, by adding a quadratic growth component, it could be easily tested whether a quadratic growth curve model fits the data significantly better than a linear one. In the present example, the linear model as shown in Figure 5 results in $\chi^2 = 6.180$ with 5 degrees of freedom ($p > .05$), indicating a good model fit. After introducing an additional quadratic growth factor the fit improves slightly ($\chi^2_{quadratic} = 4.138$) but the improvement is not significant ($\chi^2_{Diff} = \chi^2_{linear} - \chi^2_{quadratic} = 6.180 - 4.138 = 2.042$ with $df_{linear} - df_{quadratic} = 5 - 1 = 4$, $p > .05$) so that the more parsimonious linear model would be retained. In the linear model, the means of the latent intercept and the latent slope are both significant ($\alpha_0 = 1.207$, $p < .01$, $\alpha_1 = 1.153$, $p < .01$) indicating that average performance in the learning task is significantly different from zero at the first point of measurement and that people exhibit a significant mean improvement of about 1.153 units from one time point to the next. In this regard, LGCM is similar to repeated measures ANOVA, in that it shows that the means differ over time. By imposing a linear trajectory, however, we also test the shape of the overall curve, which can be well described by a straight line as suggested by the good

model fit. In addition to mean changes, there are significant interindividual differences in true initial performance indicated by the significant variance of the latent intercept ($var(\eta_0)$ = 0.284, $p < .01$). Finally, people show significant interindividual differences in their improvement over time. This was already suggested by the strong fan-spread pattern in Figure 4A and is mapped by the significant variance of the linear latent growth curve factor ($var(\eta_1)$ = 0.149, $p < .01$). As illustrated by Figure 4A and B, the mean changes, as well as the interindividual differences in initial performance and change over time, are well described by a linear latent growth curve model. As an alternative, it is also possible to permit (some of) the factor loadings to be freely estimated. This gets close to traditional factor analysis, where no predefined growth curves are imposed on the data. In the present example we could modify the linear LGCM by allowing the last two factor loadings of η_1 to be freely estimated and let the data "tell us" the best shape of the trajectory. This would fit a "linear spine" to the data (see Meredith & Tisak, 1990), resulting in the two loadings λ_{31} = 1.772 and λ_{41} = 2.767 and a model fit of χ^2 = 1.558 with 3 degrees of freedom. However, because the improvement in model fit over the more restrictive linear model is not significant (χ^2_{Diff} = 6.180 - 1.558 = 4.622, df_{Diff} = 2, $p > .05$), the more parsimonious linear LGCM should be retained. Regardless of whether the factor loadings are freely estimated or fixed based on an existing theory, the present example pointed out that LGCM combines the analysis of mean changes, as provided by traditional analysis of variance techniques, with a more detailed analysis of the within-subject covariance matrix. It is this shift in focus – away from group changes towards individual changes – which makes LGCM such a versatile and promising technique.

To summarize, it has been shown that repeated measures ANOVA and MANOVA are essentially special cases of the more general latent growth curve modeling approach. That being said, differences exist with respect to the underlying estimation procedure. LGC-models are typically based on (large-sample) ML estimation, while least square estimation is employed for (finite-sample) ANOVA and MANOVA type models. Different estimation techniques are based on different assumptions (e.g., multivariate normality and/or a sufficiently large sample) and will produce different results based on the degree to which the assumptions are met in a given sample. A more detailed comparison of different estimation techniques is beyond the scope of this article, and I settle for a comparison on the model level in the present paper. Regarding model specification, however, it has been shown that the standard repeated measures ANOVA is identical to a LGC-model with the assumption of a spherical covariance matrix among the latent growth factors. Based on research by Raykov (2001) it was demonstrated that the assumption of compound symmetry and sphericity can be easily tested within the LGCM framework and a likelihood-ratio based alternative to Mauchly's criterion *W* has been proposed. Likewise, alternative likelihood-ratio tests were proposed for polynomial contrasts, which have been demonstrated to be easily incorporated into the general latent growth curve approach. Other than repeated measures ANOVA, neither the multivariate analysis of variance, nor LGCM rests on the assumption of sphericity. As a matter of fact, the saturated LGC-model is equivalent to MANOVA, but other than MANOVA, latent growth curve modeling allows the researcher to impose specific constraints on the covariance matrix of the latent variables. Again, a likelihood-ratio test has been proposed as an alternative to the multivariate *F*-test employed in MANOVA but future research is necessary to evaluate the validity of such a test. Finally, it has been argued that LGCM is characterized by a shift in focus, away from the analysis of mean changes, towards the analysis of individual trajectories. This change in perspective is characterized by (A) the

possibility to formulate and test much more sophisticated hypotheses regarding the within-subject covariance matrix than possible with traditional methods. (B) The need to have a good theory underlying one's model specification, and (C) great flexibility of incorporating predictors of change, which will be the topic of the next section.

Predicting change

The biggest advantage of being able to better describe (individual) changes over time is the possibility to better *predict* these changes. In this last section I will demonstrate how to use categorical and continuous variables to explain interindividual differences in change. Traditional methods will be compared to the more general LGCM approach. As before, I will proceed in three steps by first considering two-wave data before moving to more complex multi-wave designs. Finally, some more recent developments will be outlined.

Predicting change in two-wave data

The prediction of pre- to post-test change is very straightforward. As demonstrated in the first section, a paired samples *t*-test corresponds to a latent growth curve model as shown in Figure 1. This model can be easily extended by regressing η_0 and/or η_1 on potential predictors as illustrated by Figure 6A. As also discussed above, the paired samples *t*-test is identical to an independent *t*-test on the difference scores (i.e., x_4 - x_1). Thus, a regression of η_1 on group membership is identical to the regression of the difference scores on group membership, as long as the error variances of x_4 and x_1 are constrained to zero. Likewise, a regression of the observed difference scores on one or more categorical and/or continuous vari-

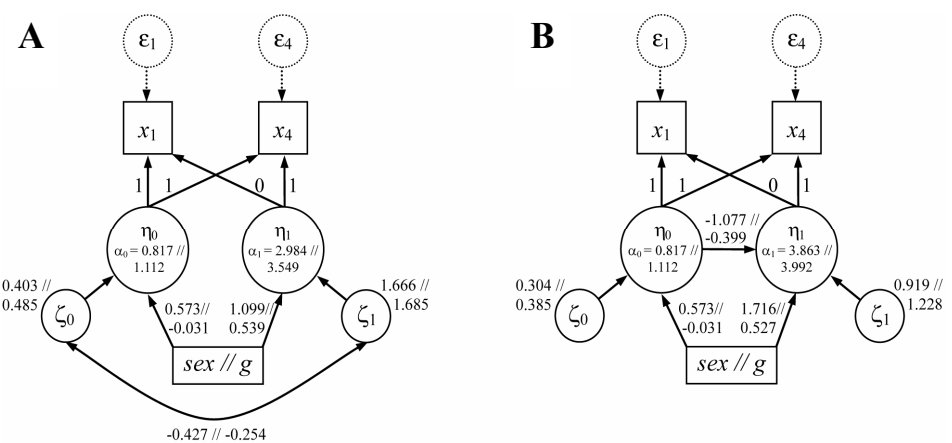

Figure 6:
Path diagram of a paired samples *t*-test (A) and a base-free measure of change model (B) using either "sex" or "g" as predictor for individual difference in pre- to post-test (x_1 to x_4) changes (sex // g). The predictor g is *z*-standardized. Men are coded as 1 and women as 0.

ables is identical to the prediction of η_1 by the same predictors. Figure 6A illustrates this fact for either sex or *g* as predictor of initial performance (η_0) and change over time (η_1). If the independent variable is in deviation form (i.e., its mean is zero), the intercept of the dependent variable corresponds to its mean (e.g., Aiken & West, 1991). For this purpose, *g* was *z*-standardized prior to including it as a predictor. Now, the mean of the latent growth factor ($\alpha_1 = 3.549$) is equal to the mean difference between pre- and post-test. As demonstrated in the first section (see Equation 5), this difference is highly significant ($p < .01$). By regressing η_1 on *g*, a regression coefficient ($\gamma_{1g} = 0.539$) is obtained, which maps the difference in improvement from pre- to post-test between people with an average intelligence (i.e., *g(standardized)* = 0) and people one standard deviation above average (*g(standardized)* = 1). Dividing the coefficient by its standard error (*SE* = 0.223), we find that the difference in mean performance is significant ($p < .05$). The same is true for the prediction of η_0, where the estimated regression coefficient ($\gamma_{0g} = -0.031$) indicates that people one standard deviation above average on *g*, start off somewhat worse at the beginning of the learning task as compared to people with an average intelligence. The difference, however, is small and not significant (*SE* = 0.119, $p > .05$). The same interpretation holds for using the categorical variable gender instead of *g* as a predictor (see again Figure 6A). In our example women are coded as zero, so the mean of η_0 corresponds to the average performance of women at the first point of measurement. The regression coefficient ($\gamma_{0sex} = 0.573$) indicates that men are slightly better than women in their initial performance and dividing γ_{0sex} by its standard error reveals that this difference is indeed significant (*SE* = 0.215, $p < .05$). Likewise, men show a larger improvement from pre- to post-test than women. On average, women improve about $\alpha_1 = 2.984$ units as compared to an increase of $\alpha_1 + \gamma_{1sex} = 2.984 + 1.099 = 4.083$ units of men. The difference is again significant ($\gamma_{1sex} = 1.099$, *SE* = 0.437, $p < .05$). The test is asymptotically equivalent to the independent samples *t*-test (using the pre-post difference scores ($x_4 - x_1$) as dependent variable) and the repeated measures ANOVA for two time points and one (categorical) between-subject factor. Table 4 compares the LGCM estimate (γ_{1sex}) with the results of a *t*-test and repeated measures ANOVA obtained by using any major statistical software package. Finally, the covariance between ζ_0 and ζ_1 maps the relationship between pre- and post-test after controlling for any predictors. Using gender as a predictor, there is a slight, but significant correlation between pre- and posttest ($\phi_{01} = -0.427$, *corr* = -0.437, *SE* = 0.156, $p < .05$), while the correlation gets smaller and is no longer significant after controlling for *g* instead of gender ($\phi_{01} = -0.254$, *corr* = -0.259, *SE* = 0.159, $p > .05$). This suggests that part of the covariation between pre- and post-test is "caused" by intelligence. However, because the sample size is small (and of course the fact that the data were chosen for illustrative rather than substantive purposes) one must be careful in interpreting this finding.

As demonstrated in the first section of this paper, the LGCM approach allows us to take the reliabilities of x_1 and x_4 into account, should they be known. As before, a reliability of $r_{tt}(x_1) = .80$ and $r_{tt}(x_4) = .85$ was assumed. Likewise, predictors can be included in the base-free measure of change model as shown in Figure 6B. Now, the effect of gender on improvement (η_1) is independent of any prior performance. That is, if men and women would have started out equal on the learning task, they would still differ in their change from pre- to post-test by about 1.716 units. The estimated difference is larger than in the previous (difference score) model and is highly significant ($\gamma_{1sex} = 1.716$, *SE* = 0.435, $p < .05$). The

Table 4:
Independent samples *t*-test, repeated measures ANOVA and LGCM approach to testing the difference in pre- to post-test improvement between men and women. Note that (asymptotically) all three approaches will yield identical results

		Independent samples *t*-test	
Mean difference (*M* vs. *F*)	*t*	*df*	*p*
1.099	-2.444	33	.020
		Repeated measures ANOVA	
Source	*F*	*df*	*p*
Time	246.979	1	.000
Time * sex	5.972	1	.020
		LGCM	
	Estimate (γ_{1sex})	*Standard error (SE)*	*p*
	1.099	0.437	.020

Note: $t^2 = F$ for $df_{numerator} = 1$; M = Male, F = Female.

effect of gender on η_0 remains unaffected by analyzing residualized (true) gain scores instead of direct difference scores. However, it is now possible to obtain and test the indirect effect of gender via η_0 on η_1. The estimate is simply computed by multiplying β_{10} with γ_{0sex}. The standard error is readily provided by most software packages (command IND for Mplus). The indirect effect ($\gamma_{0sex} * \beta_{10} = 0.573 * -1.077 = -0.617$), however, is not significant ($SE = 0.322$, $p > .05$). The same procedure can be adopted in testing the direct and indirect effects of *g* instead of sex (see Figure 6B). Note, that the LGCM approach makes no difference between categorical and continuous predictors. This stands in sharp contrast to the repeated measures ANOVA, where the between-subject factor must be categorical. If this is the case, the results are identical as shown in Table 4. ANOVA, however, cannot be employed if *g* would be used as a predictor instead of gender. In the case of two-wave data, taking the difference between x_4 and x_1 and regressing it on any continuous predictor easily circumvents this problem. This cannot be done in multi-wave data as will be discussed in the next section. In addition, the predictor is assumed to be measured without error in standard regression or ANOVA type procedures. This is also no longer true for the LGCM approach, where we can easily adjust for unreliability of the predictor in the same way the dependent variable(s) were adjusted for unreliability. Moreover, the independent variables need not be directly observed, but may be latent. This is only possible in the LGCM approach.

Of course it is possible to consider more than one predictor at a time and to include interactions among predictors. All this is not different from any standard regression analysis and shall not be reviewed in this paper (but see Cohen et al., 2003; Aiken & West, 1991). In structural equation modeling, *multiple group analysis* is another option to test hypotheses involving categorical (grouping) variables. An introduction to multiple group analysis can be found in any SEM textbook (e.g., Bollen, 1989, pp. 355). One advantage of multiple group analysis is the possibility to test for differences in variances across groups instead of being limited to mean differences as it is the case in linear regression. From the previous analyses we know that men are significantly better, both, in true initial performance and improvement

from pre- to post-test. Accordingly, we might set up a model, which accounts for this fact by allowing the means of the two latent variables to differ across groups. In a next step, however, it might be interesting to see whether men are not only better, but exhibit larger interindividual differences as compared to women. This test is readily implemented by comparing a model where the variances of η_0 and η_1 are constrained to equality across groups to a model where the variances are allowed to be freely estimated. Figure 7 shows the parameter estimates of the unconstrained model. Constraining both variances to equality ($var(\eta_{0men}) = var(\eta_{0women}) = 0.408$, $SE = 0.097$, $p < .01$ and $var(\eta_{1men}) = var(\eta_{1women}) = 1.602$, $SE = 0.382$, $p < .01$), we obtain a model fit of $\chi^2 = 7.546$ with 2 degrees of freedom. Allowing the variances to differ, the model is just identified with zero degrees of freedom, so the χ^2 reported above indicates that the two variances (taken together) differ significantly across groups ($p < .05$). As apparent from Figure 7, men exhibit an over three times larger variance at the pre-test, while the post-test variance is almost identical for men and women ($var(\eta_{0men}) = 0.627$, $SE = 0.209$, $p < .05$; $var(\eta_{1men}) = 1.646$, $SE = 0.549$, $p < .05$ and $var(\eta_{0women}) = 0.167$, $SE = 0.057$, $p < .05$; $var(\eta_{1women}) = 1.687$, $SE = 0.579$, $p < .05$). As a matter of fact, when constraining the post-test variance to equality, the drop in fit is negligible suggesting that men and women do not differ in their variability of pre- to post-test change ($\chi^2_{Diff} = 7.546 - 7.056 = 0.49$, $df_{Diff} = 2 - 1 = 1$, $p > .05$). Reintroducing *g* as a predictor of η_0 and η_1 allows us to test (possibly quite sophisticated) interaction hypotheses. For example, by comparing a model where γ_{1g} is constrained to equality across the two groups to a model where γ_{1g} is allowed to differ, is equivalent to testing an interaction between *g* and sex. A significant χ^2_{Diff} would suggest that the effect of *g* on pre- to post-test change differs between men and women. One can think of numerous other hypotheses, which are readily formulated and tested within this general

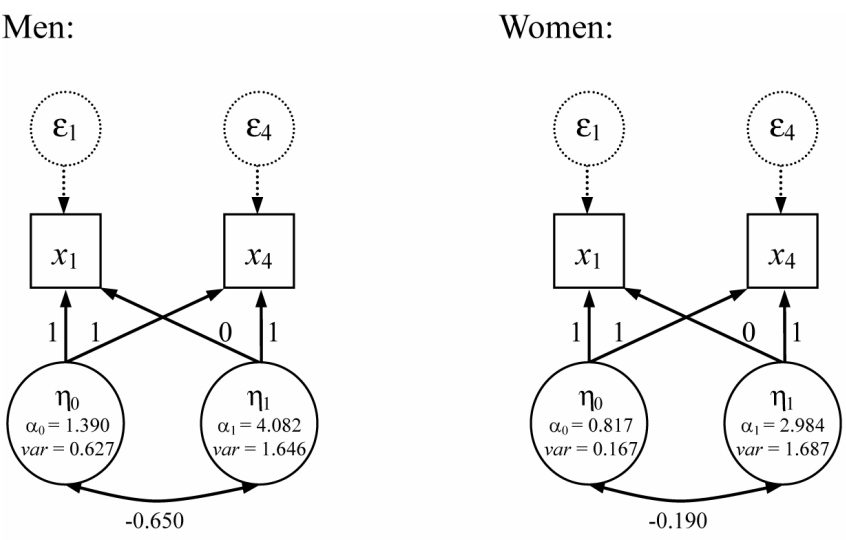

Figure 7:
Multiple group analysis using gender as grouping variable. All freely estimated parameters are allowed to differ across groups (saturated model)

framework. As mentioned in the first section of this paper, the LGCM approach gets particularly interesting if more than one (parallel) measure was obtained at each time point, so that the error terms can be estimated rather than being constrained a priori.

Predicting change in multi-wave data

The option to use difference scores as dependent variable in a pre- to post-test analysis makes it easy to analyze and predict change in two-wave data. This is no longer true for the prediction of change over multiple waves. As discussed in the previous section, traditional methods are often based on very restrictive assumptions, such as compound symmetry or sphericity, which still apply when it comes to the *prediction* of change. Furthermore, there are additional assumptions that must be met when including predictors. Finally, the exclusive focus on mean changes – instead of individual trajectories – restricts MANOVA and ANOVA to the use of categorical predictors. The more general LGCM approach is not only more flexible with respect to those assumptions, but offers a convenient way to test them. Moreover, it is not limited to categorical predictors but allows for any combination of categorical and/or continuous variables.

Using the same factor loading matrix as before,

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0.5 & -0.671 & 0.5 & -0.224 \\ 0.5 & -0.224 & -0.5 & 0.671 \\ 0.5 & 0.224 & -0.5 & -0.671 \\ 0.5 & 0.671 & 0.5 & 0.224 \end{pmatrix}$$

η_0 is again "centered", mapping average interindividual differences across all four time points. Thus, a regression of η_0 on a categorical predictor corresponds to the between-subject analysis in a repeated measures ANOVA. Figure 8 shows the according path diagram for our example data set. Constraining the regression coefficient γ_{0sex} to zero, the variance of η_0 ($var(\eta_0) = var(\zeta_0) = 2.439$) corresponds to the between-subjects sum of squares divided by n ($35 * 2.439 = 85.36 = SS_{Between}$; see Table 3). When regressing η_0 on gender ($\gamma_{0sex} = 2.230$), the remaining variance corresponds to the variance not accounted for by sex ($var(\zeta_0) = 1.197$), so that the sum of squares explained by sex are $SS_{Between}(sex) = 85.36 - 35 * 1.197 = 43.47$, corresponding exactly to the $SS_{Between}(sex)$ obtained by any standard software package. Likewise, $F = (SS_{Between}(sex) / df_{Between}) / (SS_{Between}(error) / df_{Between}(error)) = (43.47 / 1) / (41.89 / 33) = 34.25$, suggesting that there are significant differences in mean performance between men and women ($p < .01$). Since the variance of the other factors is not affected by the introduction of a predictor, the change over time can be evaluated as discussed above (compare Figure 8 and Figure 3). However, while the prediction of interindividual differences in average performance does not depend on the within-subject covariance matrix ($\boldsymbol{\Phi}$), the analysis of interindividual differences in change over time does. In addition, the introduction of a categorical predictor requires that the variance-covariance matrix ($\boldsymbol{\Phi}$) is identical across all levels of the predictor (in our example for men and women). It is important to note that the two assumptions of sphericity and homogeneity of variance are independent,

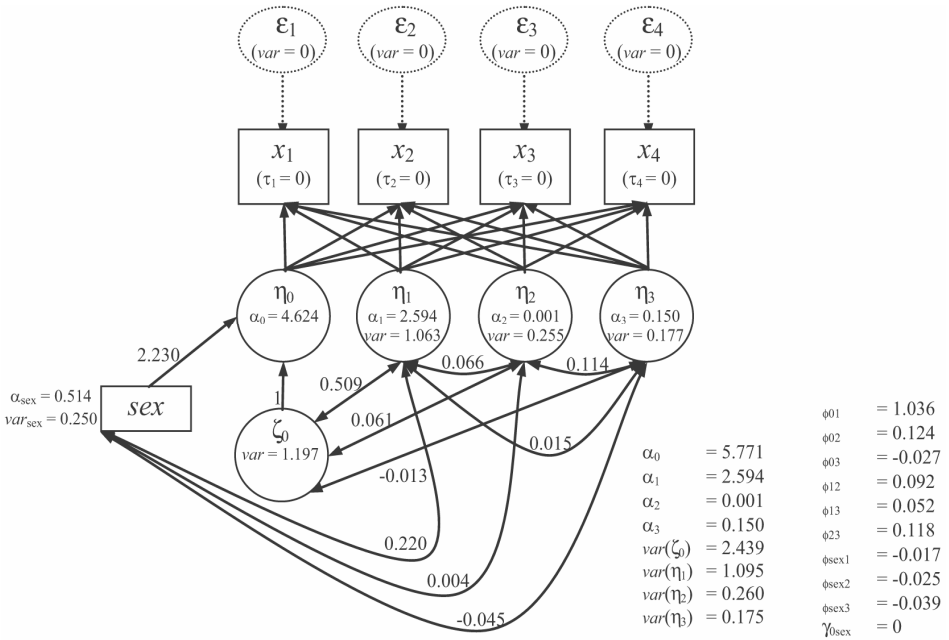

Figure 8:
Repeated measures (M)ANOVA with one between (sex) and one within (time) subject factor, where only between-subject variance is predicted by sex (no interaction between time and sex). Parameter estimates on the right side of the graph are obtained when fixing the regression coefficient of η_0 on sex to zero ($\gamma_{0sex} = 0$)

that is it can easily happen that one assumption is met, while the other is not (Maxwell & Delaney, 2000, p. 534). This makes the estimation of an interaction between a within-subject factor and a between-subject factor much more demanding. The prediction of within-subject variance is easily implemented by regressing all growth factors (η_1 - η_3) on the independent variables(s) in question. Figure 9 shows the according path diagram (with $\mathbf{\Lambda}$ as shown above). As a reminder, the within-subject main effect due to time can be computed by constraining the means of η_1 - η_3 to zero (in addition to setting $\gamma_{1sex} = \gamma_{2sex} = \gamma_{3sex} = 0$), and comparing the sum of the residual variances to the sum of residual variances after allowing the means to be freely estimated.

The *prediction* of change by gender corresponds to an interaction effect of the between-subject factor sex and the within-subject factor time. This is readily apparent from Figure 9 where the effect of sex on $\mathbf{x}$ is mediated by $\boldsymbol{\eta}$. This is not true for the prediction of the between-subject variance, because all factor loadings of η_0 are identical, thus the effect is simply multiplied by a constant. For example – and as shown above – if all factor loadings of η_0 are constrained to 0.5, the variance of η_0 corresponds to the average between-subject variance, while the mean ($\alpha_0 = 4.624$) must be divided by two in order to obtain the average performance of women on the learning task ($\bar{x}_{women} = 2.312$). When setting $\gamma_{1sex} = \gamma_{2sex} = \gamma_{3sex}$

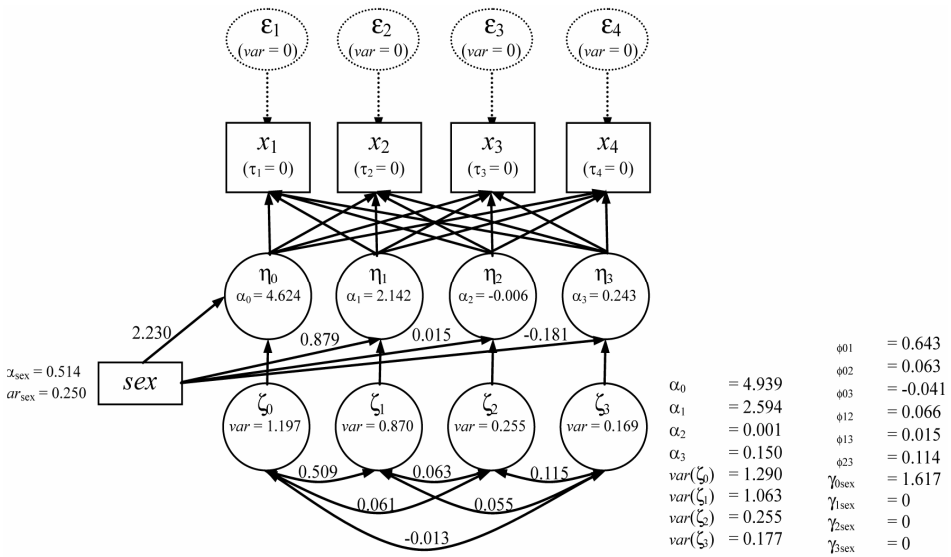

Figure 9:

Repeated measures (M)ANOVA with one between (sex) and one within (time) subject factor. Interindividual differences in change over time are predicted by sex (interaction between time and sex). Parameter estimates on the right side of the graph are obtained when setting the regression coefficients $\gamma_{1sex} = \gamma_{2sex} = \gamma_{3sex} = 0$

= 0, the sum of the variances of η_1 to η_3 is 1.063 + 0.255 + 0.177 = 1.495. Multiplying 1.495 by the number of subjects, the sum of squares between persons is $SS_{Between}$ = 35 * 1.495 = 52.325, which is the sum of squares after controlling for the within-subject factor time (note that α_1 to α_3 are freely estimated). The resulting sum of squares can be further partitioned into one part that is due to gender differences and one part that is independent of gender and independent of mean changes over time. This is readily computed by reintroducing the direct effects of sex on η_1, η_2 and η_3. The estimates are shown in Figure 9. The sum of the variances of η_1 to η_3 is 0.870 + 0.255 + 0.169 = 1.294. After multiplication with *n*, this corresponds to the $SS_{Within}(error)$ after accounting for changes over time (main effect time) and gender differences in change over time (time * sex), and is identical to the $SS_{Within}(error)$ obtained by any conventional software package ($SS_{Within}(error)$ = 45.29). As a consequence, the interaction between time and sex can be easily computed by subtracting 45.29 from 52.325, resulting in an effect of $SS_{Within}(time * sex)$ = 7.035. The *F*-test can be computed accordingly, with $F = (SS_{Within}(time * sex) / df_{Within}(time * sex)) / (SS_{Within}(error) / df_{Within}(error)) = (7.035 / 3) / (45.29 / ((35 - 2) * (4 - 1))) = 5.13$ suggesting that men and women differ significantly in their change over time ($p < .05$).

In order for the *F*-value to be a reasonable test statistic, not only the assumption of sphericity must be met, but also the assumption of equal variances and covariances across groups (i.e., the homogeneity of variance assumption). The latter is true for the traditional repeated measures approach as well as the LGCM approach. If the assumption is not met, it would not make sense to compare mean changes across groups, since the within-subject error

terms (i.e., the residual within-subject (co)variance matrix) would differ, making it a futile comparison. Other than ANOVA or MANOVA, however, LGCM provides not only a direct test of this assumption, but offers an alternative to the *F*-ratio, which neither depends on the assumption of sphericity, nor on the assumption of variance homogeneity. For this purpose, gender is not treated as an exogenous variable but as a grouping variable in a multiple-group analysis as described above. Figure 10 shows the according model. The assumption of variance homogeneity can be simply tested by comparing a model where all elements in $\boldsymbol{\Phi}$ are constrained to equality across groups[10] to a model where all elements in $\boldsymbol{\Phi}$ are allowed to differ (see Raykov, 2001). Since the two models are nested, a likelihood-ratio test can be carried out to test the significance of any differences between the models. Technically speaking, the null-hypothesis (H_0) states that $\boldsymbol{\Phi}^{(\text{men})} = \boldsymbol{\Phi}^{(\text{women})}$, while the alternative hypothesis assumes that there are significant differences between the covariance matrices of the two groups (H_1: $\boldsymbol{\Phi}^{(\text{men})} \neq \boldsymbol{\Phi}^{(\text{women})}$). For *T* repeated measures, the within-subject matrix contains $T * (T + 1) / 2$ non-redundant elements, resulting in a χ^2-difference test with $4 * 5 / 2 = 10$ degrees of freedom in our example. The test of variance homogeneity is readily implemented by constraining all ten elements in $\boldsymbol{\Phi}$ to equality across groups. The means of the latent variables are allowed to differ across groups and may be freely estimated. The constrained model results in a χ^2 of 18.262 with 10 degrees of freedom. Since this value must be compared to a saturated model with $\chi^2 = 0$, the model fit indicates that the homogeneity assumption may be met ($p > .05$). As before, the reader is reminded that the test is actually a large sample test and its performance is not very well known in small samples. In large samples, however, it might be an interesting alternative to the popular Box *M* test (Box, 1949) as pointed out by Raykov (2001).

$$M = (N - G)\log\left|\boldsymbol{S}\right| - \sum_{g=1}^{G}(N_g - 1)\log\left|\boldsymbol{S_g}\right| \tag{17}$$

As shown in Equation (17), Box's *M* statistic is also based on the likelihood-ratio test. *G* is the number of groups ($g = 1 \ldots G$; e.g., males and females), N_g is the sample size in each group, and $\boldsymbol{S_g}$ the within group covariance matrix. **S** is the covariance matrix pooled across all groups (i.e., $\boldsymbol{S} = \sum_{g=1}^{G}(N_g - 1)\boldsymbol{S}_g / (N - g)$). In the present example, $M = 17.183$. For small samples an *F* approximation is used to compute its significance, indicating that the covariance matrices are not significantly different across groups ($F = 1.491$, $df_1 = 10$, $df_2 = 5163.441$, $p > .05$; see Box (1949) for details). This stands in contrast to the assumption of sphericity, which was clearly violated. Notice that sphericity was not tested by either of the two tests, although a combined test of sphericity and variance homogeneity would be possible using the LGCM approach. With respect to variance homogeneity, the LGCM based test and Box's test yield very similar results in our example. However, since the LGCM approach requires large samples, Box's *M* statistic may be better suited for small sample sizes. Having said that, Box's test appears to be overly sensitive to non-normality (Tabachnick &

[10] Constraining all elements to equality across groups results in the same parameter estimates as shown in Figure 3. However, because the factor loadings of η_0 are now constrained to 0.5 (instead of 1.0 as in Figure 3), all estimates associated with η_0 in Figure 3 must be multiplied by two in order to obtain the same estimates.

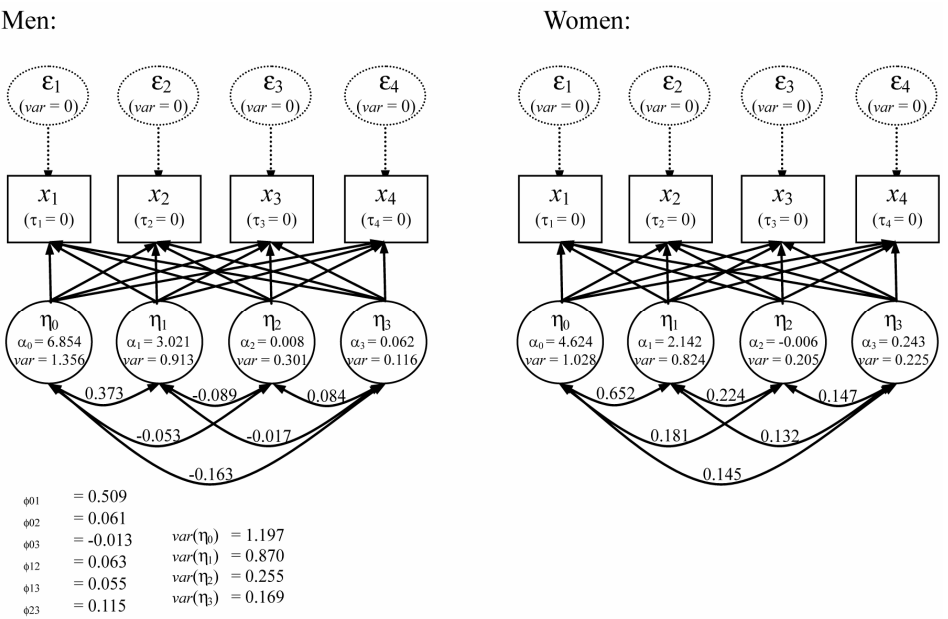

Figure 10:

Multiple group analysis using gender as grouping variable. All freely estimated parameters shown in the path diagram are allowed to differ between groups (saturated model). Parameter estimates on the left side of the graph are obtained when constraining the means of the four factors to equality across groups (not within groups). The assumption of variance homogeneity is tested by comparing the two models

Fidell, 2001; Stevens, 2002), a problem that might be less severe for the LGCM based test with large sample sizes. Future research might address this apparent trade-off by means of a Monte Carlo Simulation.

The LGCM approach offers not only a direct test of the assumptions of sphericity and/or homogeneity of variance, but can account for these violations. The multiple group analysis allows researchers to formulate and implement very specific hypothesis regarding possible group differences in **Φ**. Given the usual assumptions of structural equation models are met (primarily multivariate normality and/or a sufficiently large sample size), the resulting parameter estimates and significance tests (i.e., likelihood-ratio tests) are correct. Note that this is not true for the *F*-ratio as described above, which always depends on the assumption of variance homogeneity across groups. Another important advantage of LGCM is the option to use continuous predictors instead of categorical predictors. This is not possible in standard ANOVA or MANOVA and opens up a wide field of new applications. In our example this would simply mean replacing sex by *g* in Figure 8 and Figure 9. Finally, it is straightforward to combine the use of continuous and/or categorical exogenous predictors and multiple group analysis (i.e., categorical predictor). This is a great improvement over traditional techniques, because it allows a detailed analysis of interindividual differences in individual change over time making use of all available information (e.g., no need for categorization of continuous

predictors). At the same time, the assumptions of variance homogeneity and sphericity are no longer indispensable, but may be relaxed if necessary. However, future research is necessary to explore the differences of the approaches with respect to accuracy, power and robustness under various conditions. The great flexibility of LGCM is certainly an advantage, but also calls for a thorough matching of theory and statistical modeling. For example if change over time can be described by a linear function as demonstrated in the previous section of the present paper (Figure 5), it is easy to explain interindividual differences in initial performance as well as interindividual differences in intraindividual change over time by the introduction of level two predictors. This is illustrated in Figure 11, where gender is used to predict η_0 and η_1. In this example, men and women differ significantly in initial performance ($\gamma_{0sex} = 0.467$, $p < .05$) as well as linear change over time ($\gamma_{1sex} = 0.399$, $p < .05$). The use of additional categorical and/or continuous predictors (such as g) is straightforward. Numerous examples of conditional latent growth curve models can be found in the literature (e.g., Bollen & Curran, 2006; Duncan, Duncan, & Strycker, 2006). Of course the use of level two predictors is contingent on an adequate description of the change process, so that eventually it is up to the researcher to define the most appropriate model for his or her purposes.

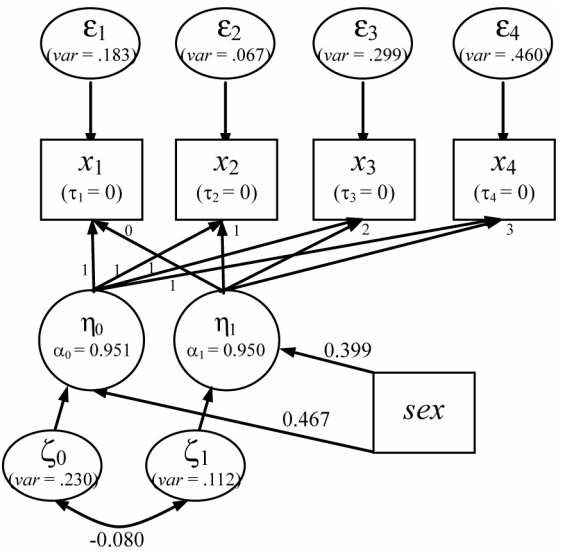

Figure 11:
Conditional linear latent growth curve model using gender as level two predictor

Extensions

Throughout the last years, the basic latent growth curve model has been extended in numerous ways. An overview of these extensions is beyond the scope of this article, but is provided for example by Bollen and Curran (2006), Duncan, Duncan and Strycker (2006), or in the excellent book edited by Moskowitz and Hershberger (2002). With respect to the prediction of change, however, there are at least three extensions worth mentioning. First, the effect of any time-invariant predictor on **x** must not necessarily be mediated by the growth factors but may be direct and time-varying (Stoel, van den Wittenboer, & Hox,

2004). This is easily accomplished by regressing **x** directly on the predictor(s) with all regression weights allowed to be freely estimated. Second, the predictors themselves can change over time, what is not possible in traditional repeated measures ANOVA or MANOVA designs (e.g., Bollen & Curran, 2006, pp. 192). The impact of time-varying predictors may again change over time or be time-invariant (i.e., the regression weights are constrained to equality). Third, it is possible to estimate parallel growth processes with average performance and/or rate of change of one process affecting average performance and/or rate of change on another variable. An introduction is given for example by Bollen and Curran (2006, pp. 198) or Curran and Willoughby (2003). Finally, the LGCM approach has been combined with other techniques to analyze change over time such as the autoregressive model (Bollen & Curran, 2004). All of these extensions are not possible with traditional techniques such as the paired samples *t*-test, ANOVA or MANOVA, making a comparison impossible. For more detailed information on these extensions, the interested reader is referred to the above-mentioned literature.

To summarize, it has been shown that LGCM is a very general approach to the prediction of change. The conventional *t*-test, repeated measures ANOVA and MANOVA – with and without between-subject and within-subject factor – are all special cases of the more general latent growth curve approach. For the most simple case of only two time points, the use of difference scores constitutes an easy way to analyze interindividual differences in change. In such a case, and for only two groups, the independent samples *t*-test, repeated measures ANOVA with one between-subject factor, MANOVA and LGCM as shown in Figure 6A yield identical results. In addition, the LGCM approach offers a convenient way to analyze and predict residualized (true) gain scores. Finally, LGCM can account for imperfect reliability of the criterion and/or predictor. Especially for more complex models with multiple measurements taken at each time point and complex (latent) predictors, this is a great improvement over traditional techniques. For multi-wave data it has been shown that the repeated measures design with one between- and one within-subject factor (also known as split-plot or mixed design, see Maxwell & Delaney, 2000, pp. 517) can be easily incorporated into the more general LGCM approach. Other than the traditional methods, however, LGCM is not limited to the use of categorical predictors. As a matter of fact, quite complex interaction hypotheses including categorical and/or continuous predictors can be tested. The assumption of variance homogeneity across groups, which is crucial for repeated measures ANOVA and MANOVA, can be tested within the LGCM framework, but other than ANOVA or MANOVA, LGCM can also account for violations of this assumption by offering an alternative to the conventional *F*-test. The most striking difference between LGCM and ANOVA/MANOVA is the greater flexibility of the former as compared to the latter. While this is certainly an advantage, it demands great diligence from the researcher when setting up the model and interpreting results.

Discussion

By demonstrating that the analysis of variance and multiple regression are essentially identical data analytic systems, Cohen (1968) prepared the ground for a new way of statistical thinking among social scientists. Instead of treating ANOVA and multiple regression as different techniques, he pointed to the generality of MR, which comprises the analysis of

variance as a special case. This prepared the ground for more refined analyses regarding group differences and interindividual differences, helping ultimately to bridge the gap between experimental and differential psychology. In the present paper I have argued that it is time for a similar reconceptualization in the analysis of change. During the last decade there has been an almost exponential increase in methodological and applied articles using "new methods for the analysis of change" (Collins & Sayer, 2001). The new procedures focus on intraindividual variability instead of mean changes, which have been of central interest in traditional methods such as the paired samples *t*-test, repeated measures ANOVA or MANOVA. The notion that the latter are just a special case of the former has always been present (Meredith & Tisak, 1990), but most of the current literature treats techniques rooted in the analysis of variance (i.e., *t*-test ANOVA, MANOVA) and factor analytic techniques (i.e., latent growth curve modeling) as largely unrelated. I think that this is unfortunate, because much can be learned about either approach by examining their commonalities as well as their differences. Of course there exist some noteworthy exceptions, for example chapter three in Duncan, Duncan and Strycker (2006) to name just one, but I am not aware of any comprehensive treatment of this topic. The present paper attempts to fill this gap in a didactic manner by demonstrating the equivalence of traditional techniques (*t*-test, base-free measures of change, repeated measures ANOVA, polynomial contrasts, MANOVA) and the more general latent growth curve models, if certain assumptions are met and certain constraints are imposed. All arguments have been illustrated by a hypothetical data set on skill-acquisition.

There are a number of problems associated with such a didactic approach. First and foremost, it falls short to set out the mathematical relationship between the models introduced, even in cases where it would be possible to do so. In addition, the relationship between models can only be demonstrated on a conceptual (model-) level, since the actual estimates are affected by the estimation procedure, rounding errors and even differences in the software packages employed – although the last two issues are largely negligible. Especially the estimation procedure, however, depends greatly on sample size. Thus finite-sample differences in parameter estimates can be quite substantial. This is especially true if the sample is as small as the one used in the present paper. On the other hand, the use of a small set of raw data, enables the reader to reproduce all analyses and results using different software packages or even hand calculation where possible. In addition, it is much easier to follow a simple, albeit artificial, example as compared to the typically much more complex real-world studies. Because of the didactic nature of this paper I opted for the small sample example.

Considering the fact that the sample size in the present example is clearly too small for the more complex (e.g., multiple group) LGCM analyses, it is surprising that most parameter estimates turned out to be quite similar to the ones obtained by traditional methods in our example. Nevertheless, it must be emphasized – once again – that LGCM is a large sample method and cannot be recommended if the sample size is small and the assumptions of traditional methods are met. Having said that, it is difficult to tell when the sample size will be "too small" for LGCM. As long as the analysis is restricted to simple (saturated) models focusing on mean changes, LGCM and traditional techniques yield identical results even for very small samples. For more complex models, however, differences can be quite substantial. On the other hand, obvious violations of central assumptions underlying traditional techniques (such as sphericity and homogeneity of variance) may justify the use of LGCM

despite an "insufficient" sample size. Clearly, there is a need for future research to shed light on the complex interaction between these factors (sample size, underlying assumptions, model complexity) in order to determine the optimal procedure for the analysis of change for a given set of data.

Similar arguments can be made for most fit indices employed in LGCM, which are also greatly affected by sample size. This topic has been deliberately ignored because it is no different from standard structural equation modeling and a more detailed discussion would go far beyond the scope of this article.

If sample size is sufficiently large, LGCM can be conceived as a general data analytic approach to the analysis of change. As discussed throughout the paper, it comprises many traditional methods as special cases. It offers direct tests of important assumptions and allows researchers to account for potential violations of these assumptions. In addition, it can easily handle categorical as well as continuous variables. Its biggest advantage over conventional techniques such as the *t*- test, ANOVA or MANOVA, however, is its flexibility with respect to the specification of the within-subject variance-covariance matrix **Φ**. Almost any hypothesis regarding interindividual differences in intraindividual change can be tested by imposing specific constraints on **Φ**. This argument also generalizes to the prediction of change as discussed in the last section of the article. It is this shift in focus – and no substantial differences – that make "new methods for the analysis of change" different from "traditional" techniques. At the same time, however, it shows how the approaches relate to each other and how they can be integrated.

It is hoped that the present article appeals to the applied and methodologically interested reader alike. Several avenues for future methodological research have been pointed out, mainly relating to power, precision and robustness of the different approaches. At the same time applied researcher are encouraged to pay more attention to the specification of the within-subject covariance matrix. Oftentimes, it is possible to formulate quite specific research questions and LGCM offers the flexibility to address these questions.

References

Ackerman, P. L. (1992). Predicting Individual Differences in Complex Skill Acquisition: Dynamics of Ability Determinants. *Journal of Applied Psychology, 77*(5), 598-614.

Ackerman, P. L., & Kanfer, R. (1993). Integrating Laboratory and Field Study for Improving Selection: Development of a Battery for Predicting Air Traffic Controller Success. *Journal of Applied Psychology, 78*(3), 413-432.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications.

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

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The Role of Coding Time in Estimating and Interpreting Growth Curve Models. *Psychological Methods, 9*(1), 30-52.

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

Bollen, K. A., & Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) Models: A Synthesis of Two Traditions. *Sociological Methods & Research, 32*, 336-383.

Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective*. Hoboken, NJ: John Wiley.

Bollen, K. A., & Long, S. J. (Eds.). (1993). *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Box, G. E. P. (1949). A General Distribution Theory for a Class of Likelihood Criteria. *Biometrika, 36*(3/4), 317-346.

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

Campbell, D. T., & Erlebacher, A. (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Compensatory education: A national debate* (Vol. 3). New York: Brunner/Mazel.

Cattell, R. B. (Ed.). (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.

Cohen, J. (1968). Multiple Regression as a General Data-Analytic System. *Psychological Bulletin, 70*(6), 426-443.

Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Hillsdale: Erlbaum.

Collins, L. M., & Sayer, A. G. (Eds.). (2001). *New Methods for the Analysis of Change*. Washington DC: American Psychological Association.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.

Cronbach, L. J. (1975). Beyond the Two Disciplines of Psychology. *American Psychologist, 30*, 116-127.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"-or should we? *Psychological Bulletin, 74*(1), 68-80.

Curran, P. J., & Willoughby, M. (2003). Implications of latent trajectory models for the study of developmental psychopathology. *Development and Psychopathology, 15*, 581-612.

DuBois, P. H. (1957). *Multivariate correlational analysis*. New York: Harper.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications:* (Second ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Greenhouse, S. W., & Geisser, S. (1959). On Methods in the Analysis of Profile Data. *Psychometrika, 24*(2).

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Huynh, H., & Feldt, L. S. (1970). Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions. *Journal of the American Statistical Association, 65*(332), 1582-1589.

Kenny, D. A. (1974). A quasi-experimental approach to assessing treatment effects in nonequivalent control group designs. *Psychological Bulletin, 82*, 345-362.

Keselman, H. J., & Rogan, J. C. (1980). Repeated measures F tests and psychophysiological research: Controlling the number of false positives. *Psychophysiology, 17*, 499-503.

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. (1980). Testing the Validity Conditions of Repeated Measures F Tests. *Psychological Bulletin, 87*(3), 479-481.

Levy, R., & Hancock, G. R. (2007). A framework of statistical tests for comparing mean and covariance structure models. *Multivariate Behavioral Research* (42), 33-66.

Lohman, D. F. (1999). Minding our p's and q's: On finding relationships between learning and intelligence. In P. L. Ackerman, P. C. Kyllonen & R. D. Roberts (Eds.), *The future of learning and individual differences: Process, traits, and content* (pp. 55-72). Washington, DC: American Psychological Association.

Lord, F. M. (1956). The measurement of growth. *Educational and Psychological Measurement, 16*, 421-437.

Lord, F. M. (1963). Elementary Models for Measuring Change. In C. W. Harris (Ed.), *Problems in Measuring Change* (pp. 21-38). Madison: University of Wisconsin Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mauchly, J. W. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. *The Annals of Mathematical Statistics, 11*(2), 204-209.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

McArdle, J. J., & Nesselroade, J. R. (2002). Growth Curve Analysis in Contemporary Psychological Research. In J. Schinka & W. Velicer (Eds.), *Comprehensive Handbook of Psychology, Volume Two: Research Methods in Psychology* (pp. 447-480). New York: Wiley.

McNemar, Q. (1958). On Growth Measurement. *Educational and Psychological Measurement, 18*(1), 47-55.

Mendoza, J. L. (1980). A Significance Test for Multisample Sphericity. *Psychometrika, 45*(4), 495-498.

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

Meredith, W., & Tisak, J. (1990). Latent Curve Analysis. *Psychometrika, 55*(1), 107-122.

Moskowitz, D. S., & Hershberger, S. L. (Eds.). (2002). *Modeling Intraindividual Variability with Repeated Measures Data: Methods and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B., & Muthén, L. K. (1998-2007). *Mplus [computer software]*. Los Angeles: Muthén & Muthén.

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

Raykov, T. (1992). Structural models for studying correlates and predictors of change. *Australian Journal of Psychology, 44*(2), 102-112.

Raykov, T. (1993a). On estimating true change interrelationships with other variables. *Quality & Quantity, 27*, 353-370.

Raykov, T. (1993b). A Structural Equation Model for Measuring Residualized Change and Discerning Patterns of Growth or Decline. *Applied Psychological Measurement, 17*(1), 53-71.

Raykov, T. (1999). Are Simple Change Scores Obsolete? An Approach to Studying Correlates and Predictors of Change. *Applied Psychological Measurement, 23*, 120-126.

Raykov, T. (2001). Testing Multivariable Covariance Structure and Means Hypotheses via Structural Equation Modeling. *Structural Equation Modeling, 8*(2), 224-256.

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

Rogosa, D. R. (1995). Myths and Methods: "Myths About Longitudinal Research" plus Supplemental Questions. In J. Gottman (Ed.), *The analysis of change* (pp. 3-66). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A Growth Curve Approach to the Measurement of Change. *Psychological Bulletin, 92*(3), 726-748.

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement, 20*(4), 335-343.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*(2), 203-228.

Rovine, M. J., & Molenaar, P. C. M. (1998). The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research, 3*(2), 95-107.

Schermelleh-Engel, Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online, 8*(2), 23-74.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press, Inc.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling True Intraindividual Change: True Change as a Latent Variable. *Methods of Psychological Research Online, 2*(1), 21-33.

Stoel, R. D., & van den Wittenboer, G. (2003). Time Dependence of Growth Parameters in Latent Growth Curve Models with Time Invariant Covariates. *Methods of Psychological Research, 8*(1), 21-41.

Stoel, R. D., van den Wittenboer, G., & Hox, J. (2004). Including Time-Invariant Covariates in the Latent Growth Curve Model. *Structural Equation Modeling, 11*(2), 155-167.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn and Bacon.

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

Tucker, L. R., Damarin, F., & Messick, S. (1966). A base-free measure of change. *Psychometrika, 31*(4), 457-473.

Vasey, M. W., & Thayer, J. F. (1987). The Continuing Problem of False Positives in Repeated Measures ANOVA in Psychophysiology: A Multivariate Solution. *Psychophysiology, 24*(4), 479-486.

Wainer, H. (2000). The Centercept: An Estimable and Meaningful Regression Parameter. *Psychological Science, 11*(5), 434-436.

Willett, J. B. (1997). Measuring Change: What Individual Growth Modeling Buys You. In E. Amsel & K. A. Renninger (Eds.), *Change and Development: Issues of Theory, Method, and Application.* (pp. 213-243). Mahwah, NJ: Lawrence Erlbaum Associates.

Wishart, J. (1938). Growth rate determinations in nutrition studies with the bacon pig, and their analyses. *Biometrika, 30*, 16-28.

Wittmann, W. W. (1988). Multivariate reliability theory. Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 505-560). New York: Plenum Press.

Wittmann, W. W. (1997, July). *The reliability of change scores: Many misinterpretations of Lord and Cronbach by many others. Revisiting some basics for longitudinal research.* Paper presented at the Methodology conference: Evaluation of Change in Longitudinal Data, Nürnberg.