Methods of Psychological Research Online 2004, Vol. 9, No. 1, pp. 37-55
Internet: http://www.mpr-online.de

Decreases in Posttest Variance and The Measurement of Change

Robert A. Cribbie
York University

John Jamieson
Lakehead University

A Monte Carlo study was used to evaluate the effects of reductions in posttest variance on several methods for detecting predictors of change in a two-wave design. When the predictor was dichotomous, the analysis of covariance approach was compared to the analysis of variance on difference scores. For a continuous predictor, partial correlations, difference score correlations with the predictor and latent change correlations with the predictor in structural equation growth models were used. When posttest variance decreased (e.g., ceiling effect) difference scores lost power, while the power of regression based methods (analysis of covariance and partial correlations) and structural equation measures of change were unaffected.

In spite of arguments in favour of multiple measurements over time (e.g., Willett, 1997), the pretest-posttest design is still widely used to compare the changes exhibited by two or more groups in response to a treatment (Collins, 1996a; Williams & Zimmerman, 1996; Bonate, 2000). This design has two advantages over a posttest only design. First, the pretest provides information about individual differences, which can be used to decrease estimates of error variance, thereby increasing power. The second advantage is that baseline (pretest) differences between groups can be taken into account.

Two statistical methods are most often used to compare the changes from pretest to posttest. One method computes the posttest minus pretest difference scores (also called change or gain scores). An independent t-test or ANOVA is used to compare the mean difference scores for each group. This method is equivalent to the interaction term in a two-way mixed ANOVA, and provides a direct comparison of the mean changes exhibited by each group. The second method is analysis of covariance (ANCOVA) in which the posttest is the dependent variable and the pretest is the covariate. This method an-

Robert A. Cribbie, Department of Psychology, York University, Toronto, Canada. John Jamieson, Department of Psychology, Lakehead University, Thunder Bay, Canada.

Correspondence concerning this article should be addressed to Robert A. Cribbie, Department of Psychology, York University, Toronto, Ontario, M3J 1P3, Canada. E-mail: cribbie@yorku.ca

swers the question of whether group membership predicts differences in posttest scores after pretest differences have been removed. Both methods achieve the dual goals of removing variability due to individual differences and adjusting for baseline differences between the groups.

Difference scores and ANCOVA test different hypotheses, and can produce quite different answers (Lord, 1967). ANCOVA tests the hypothesis of equivalence of adjusted means, assuming the groups were equivalent on the pretest. Since differences in pretest means are assumed to be due to chance, a regression line is used to adjust posttest means to take into account the expected dissipation of these pretest differences at posttest (regression to the mean). This hypothesis makes sense for randomized experiments but not for quasi-experimental designs. In particular, ANCOVA should not be used to control for large baseline differences between naturally occurring groups (see e.g., Miller & Chapman, 2001), since it is not reasonable to assume the groups have the same true baseline. ANOVA on difference scores tests the hypothesis of equivalence of means of differences, regardless of the pretest differences between groups. The difference scores are assumed to yield unbiased estimates of a treatment effect which is additive and independent of the pretest level. While difference scores were spurned for many years because of concerns raised by Cronbach and Furby (1970) and others about their unreliability, more recent work has shown that reliability is not a serious problem (Llabre, Spitzer, Saab, Ironson & Schneiderman, 1991; Williams & Zimmerman, 1996). For example, Rogosa and Willett (1983) pointed out that the reliability of change scores increases whenever the true change varies across individuals.

There has been considerable uncertainty about when to use these two methods (e.g., Wainer, 1991; Maris, 1998). At present there is an emerging consensus to avoid ANCOVA for comparing changes in quasi-experimental designs (e.g., Rogosa, 1988; Schafer, 1992; Cribbie & Jamieson, 2000; Miller & Chapman, 2001). When large, naturally occurring baseline differences are present, ANCOVA is the wrong model and will produce a systematic bias that favours the finding of greater change in the group with the higher pretest mean (Jamieson, 1999). For randomized experiments ANCOVA is generally preferred to difference scores since it provides a slight increase in power (e.g., Bonate, 2000).

There has also been interest in how ANCOVA and difference scores perform under conditions in which distribution assumptions such as normality and homogeneity of variance do not hold. A recent book by Bonate (2000) presents an extensive series of computer simulations which compare the Type I error rates and power of ANCOVA and difference scores (as well as other methods) in the pretest-posttest design. He found that ANCOVA was slightly more powerful than difference scores over a wide range of distributions. However, the power of both tests decreased with departures from normality. Bonate examined several non-normal (skewed) distributions and failed to detect major differences between difference scores and ANCOVA. He also explored the effect of increasing posttest variance and concluded it had little effect on power of either method.

While Bonate studied a wide range of conditions, he failed to examine the effect of decreasing posttest variance. This was an unfortunate omission, since decreases in posttest variance are frequently found in research when a floor or ceiling effect is present.

A floor or ceiling effect arises when scores approach maximum (ceiling) or minimum (floor) values. For example, in comparing how two groups respond to a relaxation method, the posttest scores for both groups may approach a floor of zero, resulting in decreased variability. It has been known for some time that difference scores do not perform well when floor/ceiling effects are present (Jamieson, 1995). For example, Collins (1996b) stated: "assuming there are no ceiling or floor effects, there is nothing unsound about difference scores" (p. 39).

In addition to floor/ceiling effects, a decrease in posttest variance can result from a skewed measurement. With skewed measures, the means and variances are proportional (e.g., Winer, 1971) so that changes in the mean will be accompanied by changes in variance. With positively skewed measures, a mean decrease will be accompanied by a decrease in posttest variance. For example, response latency has a positively skewed distribution, and a treatment which decreases response latency will cause decreased variability in the posttest scores. Because there is a floor of zero for response latency, this is also an example of a floor effect. Bonate studied the effect of skewness, but not when it was accompanied by a change in the means, which would produce a change in posttest variance.

While difference scores and ANCOVA are the most widely used methods to compare changes in the pretest-posttest design, there has also been recent interest in structural equation models (SEM) for identifying correlates of change. Raykov (1994) and others have proposed models that include a latent variable to measure the change from pretest to posttest, and Duncan, Duncan, Strycker, Li and Alpert (1999) provide an excellent introduction to the topic. Assuming that the expected change is linear in form these models can be very useful in testing complex hypotheses regarding change and also allow the incorporation of congeneric (variables related to the underlying factor) measurements of change. The path (parameter) between latent change and a (measured or latent) predictor variable is of interest to a researcher exploring predictors of change. To date there have been no studies examining how well SEM models of change are able to detect correlates of change when posttest variability decreases.

The purpose of the present study was to compare the power of change scores, regression based methods (ANCOVA, partial correlation) and structural equation models of change, when variance decreases from pretest to posttest. Computer simulations produced a decrease in variance from pretest to posttest due to either a floor/ceiling effect, or as the result of skewness combined with a change in the means. The first set of simulations represented a two group pretest-posttest design. The power of ANOVA on difference scores and the power of ANCOVA to detect a difference in mean change between the two groups were compared. The second set of simulations replaced the two groups with a continuous predictor, with two congeneric variables used as measures of the pretest, posttest and predictor constructs. Power to detect a relationship between the predictor and change was compared for three measures: (1) the correlation of the predictor with difference scores (which is the generalization of ANOVA on difference scores); (2) the partial correlation of the predictor with posttest, controlling for pretest (which is equivalent to multiple regression and is a generalization of ANCOVA); and (3) the correlation/covariance between the latent change and the latent predictor in the structural model.

Method

The SAS generator RANNOR (SAS Institute, 1990) was used to generate pseudorandom variates, with means of zero and standard deviations (see below) selected to yield realistic effects. Two values were selected for the standard deviation of the error terms, 4 and 8, which yielded values of reliability (for the pretest) of .86 and .61, respectively. Sample sizes were selected to be representative of the designs investigated. One thousand simulations were computed for each condition.

For comparison of changes between two groups, pretest and posttest scores were created for two groups, each of $n = 25$, by adding a different error component ($\mu = 0$, $\sigma = 4$ or 8) to the same true score component ($\mu = 0$, $\sigma = 10$). Five points were added to the posttest scores of one group to create a differential change signal to be detected.

For the continuous predictor, two congeneric pretest and posttest measures were created for $N = 200$ cases by adding a different error component ($\mu = 0$, $\sigma = 4$ or 8) to the same true score component ($\mu = 0$, $\sigma = 10$). A change component ($\mu = 0$, $\sigma = 10$) was added to the posttest scores. Two congeneric predictor variables were created by adding a true score component ($\mu = 0$, $\sigma = 10$) to an error component ($\mu = 0$, $\sigma = 4$ or 8) and by adding the change component, premultiplied by .2 to create empirically realistic correlations between the predictor variables and change.

To simulate a floor/ceiling effect an increment was added to the posttest scores which was a function of the distance of the posttest scores from an arbitrary ceiling of 50. Thus, the posttest score became posttest + $c \cdot (50 - \text{posttest})$, where c had values of 0, .1, .2, .3, .4, and .5. Zero represented no ceiling effect, while .5 represented a strong ceiling effect.

To simulate skewed data, a constant of 100 was first added to all the scores to eliminate negative values. Then a constant was added to the posttest scores to simulate a mean change. The constants examined were 0, 25, 50, 75 and 100. The larger numbers represented greater change away from the tail of the distribution, while 0 represented no mean change. Changes away from the tail of the distribution are conceptually similar to a floor or ceiling effect. Finally the pretest and posttest scores were transformed to log(10) to create negatively skewed distributions.

Difference scores for the repeatedly assessed variables were computed by taking the difference between the pretest and posttest scores. ANOVA was used to compare the mean difference scores of the two groups. ANCOVA with posttest as the dependent variable and pretest as the covariate was also used to compare the changes between the two groups. Power was recorded as the proportion of simulations in which the groups were significantly different (two-tailed $\alpha = .05$).

For the continuous predictors, correlations were computed between the difference scores and each of the congeneric measures of the predictor variable, and partial correlations were computed between the posttest scores and the measures of the predictor variable, after controlling for the corresponding pretest scores. Power was recorded as the proportion of simulations in which the correlations were statistically significantly (two-tailed $\alpha = .05$).

Structural equation model

The structural model used in this study (see Figure 1) was derived from models developed by Raykov (1997), MacCallum, Kim, Malarkey and Kiecolt-Glaser (1997), Steyer, Eid, and Schwenkmezger (1997), and others, for measuring change and identifying correlates of change with SEM. The observed variables in this model are the two pretest measures, the two posttest measures, and the two measures of the predictor variable. Latent measures in this model represent baseline, change from pretest to posttest, and a predictor variable. The variances of the latent baseline, change and third variable constructs were estimated from the data. Models often include the constraint that pretest and posttest variances are equal, to reflect consistency in the same variables when measured twice; however, since the present study involved manipulations that affected posttest variance those constraints were not included. The model contains 9 degrees of freedom, with 12 unknowns being estimated from 15 covariances and 6 variances. The structural models were tested against the data using SAS PROC CALIS (SAS Institute, 1990b) with maximum likelihood estimation.

The covariance between the latent predictor and change variables was used as the measure of the relationship between the predictor variable and change. Power was recorded as the proportion of simulations in which the covariances were statistically significant (two-tailed $\alpha = .05$). As well, to allow for comparisons with the ANCOVA and difference score approaches, the correlation between the latent predictor and change variables was included. To test the fit of the model to the data, the Goodness of Fit Index (GFI) (Jöreskog & Sörbom, 1989), Adjusted Goodness of Fit Index (AGFI) (Jöreskog & Sörbom, 1989), Comparative Fit Index (CFI) (Bentler, 1988), and Root Mean Square Error of Approximation (RMSEA) (Browne & Cudeck, 1993) were recorded for each analysis, to provide a good overall picture of the fit of the model to the data.

Results

Dichotomous Predictor

The means and standard deviations of the posttest scores for the ceiling effect are presented in Table 1 (the pretest conditions were identical, and average values across conditions are also presented in the table). It can be seen that the standard deviations decrease as the ceiling is approached.

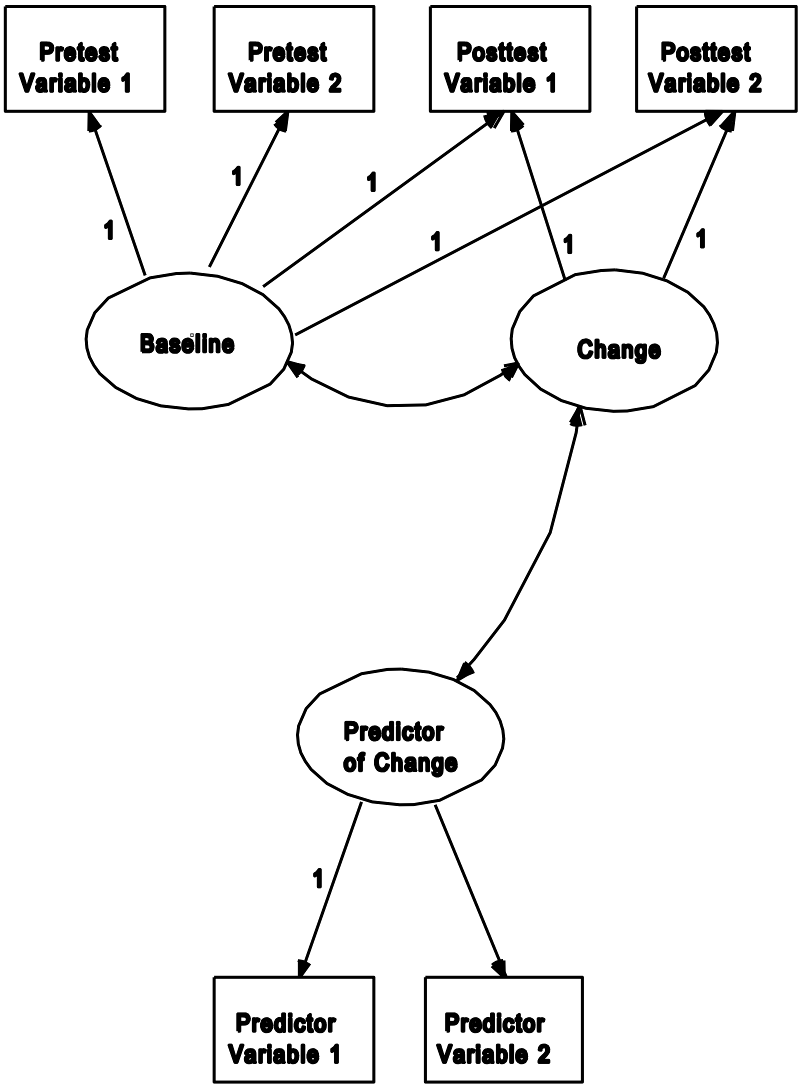

Figure 1. The structural model used for assessing the relationship between the latent predictor of change and the latent measure of change.

The proportion of simulations in which the groups were significantly different (observed power) for the ANOVA on difference scores and the ANCOVA under each ceiling effect condition are presented in Table 2. When there is no ceiling effect (change = 0), both methods have similar power to detect differences in change between the two groups. As the size of the ceiling effect increased (change toward .5), the average power for the ANOVA on difference scores decreased. In contrast, the average power for ANCOVA is unaffected by the ceiling effect.

Table 1

Average Group Means and Standard Deviations When Posttest Scores Approach a Ceiling (Group 1 Had a Constant of 5 Added to Posttest)

	Reliability (of pretest)							
	.86				.61			
	Group 1		Group 2		Group 1		Group 2	
Change	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0	5.01	10.74	0.02	10.77	5.00	12.80	0.05	12.81
.1	9.45	9.72	4.95	9.71	9.52	11.55	4.96	11.56
.2	13.98	8.66	10.01	8.67	13.94	10.24	9.93	10.23
.3	18.50	7.53	15.02	7.57	18.51	8.95	15.04	8.96
.4	22.99	6.47	19.98	6.47	22.99	7.71	20.00	7.68
.5	27.51	5.38	25.03	5.37	27.47	6.41	25.03	6.38

Note. *M* = Mean; *SD* = Standard deviation.

The means, standard deviations and skewness of the posttest scores for the negatively skewed simulations are presented in Table 3 (the pretest conditions were again identical, and average values across conditions are also presented in the table). It can be seen that the standard deviations decrease as the posttest approaches the head of the skewed distribution. As well, the skewness statistic is consistently negative, reflecting the negative skewness produced by the logarithmic transformation.

The proportion of simulations in which the groups were significantly different (observed power) for the ANOVA on difference scores and the ANCOVA for the negatively skewed simulations are presented in Table 4. The level of change indicates the amount of change in posttest scores of a negatively skewed distribution in the direction of decreasing variance (toward the head of the distribution). For the negatively skewed distribution, when there is no mean change both methods have similar average power to detect the difference in change between the two groups. However, when the amount of change away from the tail of the distribution (in the direction of decreased posttest variance) increased, the average power for the ANOVA on difference scores decreased.

As with the ceiling effect, the average power for ANCOVA was unaffected by the skewness and mean change.

Table 2

Average Power Rates (Proportion Significant) for ANOVA on Gain Scores and ANCOVA When Scores Approach a Ceiling

	Reliability (of pretest)			
	.86		.61	
Change*	Gain Scores	ANCOVA	Gain Scores	ANCOVA
0	.850	.879	.310	.371
.1	.824	.902	.280	.369
.2	.726	.908	.278	.413
.3	.568	.890	.223	.402
.4	.374	.909	.157	.382
.5	.164	.907	.097	.411

Note. * 0 = no change, .5 = half the distance of the ceiling.

Continuous Predictor

Means, standard deviations and skewness were not presented for the continuous predictor since the data are very similar to those from the dichotomous predictor. The proportion of simulations in which the correlation between the third variable and change were statistically significant (observed power) for the gain scores, partial r and SEM under each of the ceiling effect conditions are presented in Table 5. The correlation between gain scores and the predictor shows a decrease in power as the ceiling was approached. In contrast, the power of the partial correlation of the predictor with posttest (controlling for pretest), and the power of the correlation between the latent change variable and the latent predictor variable from the structural model did not decrease. Table 6 contains the average correlation between the gain scores and the predictor variable, the average partial correlation of the predictor variable with posttest (controlling for pretest), and the average correlation between the latent change variable and the latent predictor variable from the structural model. When there is no ceiling effect, all three measures yield very similar coefficients. As the ceiling effect increases, the value of the correlation between the predictor variable and the difference scores decreased, as did the SEM correlation between the predictor and change. In contrast, the partial correlations were unaffected by the ceiling effect. Although the SEM correlation between the predictor and change decreased, the power of the significance test was unaffected.

Table 3

Average Means, Standard Deviations and Skewness as Scores Change Toward the Head of a Negatively Skewed Distribution (Group 1 had a Constant of 5 Added to Posttest)

			Reliability (of pretest) = .86			
			Pretest: *M* = 1.996, *SD* = .049, skewness = -.35			
	Group 1			Group 2		
Change	*M*	*SD*	Skewness	*M*	*SD*	Skewness
0	2.019	.047	-.32	1.997	.047	-.33
25	2.112	.036	-.25	2.095	.038	-.25
50	2.189	.030	-.22	2.175	.031	-.23
75	2.254	.026	-.18	2.242	.027	-.17
100	2.311	.023	-.16	2.300	.023	-.16
			Reliability (of pretest) = .61			
			Pretest: *M* = 1.999, *SD* = .055, skewness = -.41			
	Group 1			Group 2		
Change	*M*	*SD*	Skewness	*M*	*SD*	Skewness
0	2.018	.054	-.39	1.996	.057	-.41
25	2.112	.043	-.31	2.095	.045	-.32
50	2.188	.037	-.26	2.174	.037	-.26
75	2.254	.031	-.21	2.241	.032	-.22
100	2.311	.027	-.20	2.300	.028	-.20

Note. *M* = Mean; *SD* = Standard deviation.

Table 4

Average Power Rates (Proportion Significant) for ANOVA on Gain Scores and ANCOVA When Scores Change Toward the Head of a Negatively Skewed Distribution

	Reliability (of pretest)			
	.86 (Error $SD = 4$)		.61 (Error $SD = 8$)	
Change*	Gain Scores	ANCOVA	Gain Scores	ANCOVA
0	.857	.899	.326	.399
25	.688	.879	.247	.387
50	.461	.898	.182	.399
75	.274	.872	.142	.399
100	.139	.896	.111	.411

The proportion of simulations in which the correlation between the predictor variable and change were statistically significant (observed power) for the gain scores, partial r and SEM for the negatively skewed simulations are presented in Table 7. The level of change indicates the amount of change in posttest scores of a negatively skewed distribution in the direction of decreasing variance (toward the head of the distribution). The correlation between difference scores and predictor shows a decrease in power with greater decreases in posttest variance. In contrast, the power of the partial correlation of the predictor with posttest (controlling for pretest), and of the correlation between the latent change variable and the latent predictor variable from the structural model, did not decrease.

Table 8 contains the average correlation between difference scores and the predictor variable, the average partial correlation of the predictor variable with posttest (controlling for pretest), and the average correlation between the latent change variable and the latent predictor variable from the structural model for each of the level of change conditions. When there was no mean change all three methods had similar coefficients.

However, when the amount of change toward the head of the skewed distribution increased, the value of the correlation between the predictor variable and the difference scores decreased, as did the SEM regression coefficient from the predictor to change. As with the floor/ceiling effect, the change scores but not the SEM exhibited an associated loss of power for detecting this relationship. Both the power and the value of the partial correlations were unaffected by the skewness and mean change.

Table 5

Average Power Rates (Proportion Significant) for Gain Scores, Partial r and SEM for Measuring the Correlation Between a Third Variable and Change, When Scores Approach a Ceiling

	Reliability (of pretest)					
		.86			.61	
Change*	Gain Scores	Partial r	SEM r	Gain Scores	Partial r	SEM r
0	.636	.639	.692	.240	.292	.411
.1	.629	.645	.686	.224	.307	.428
.2	.568	.630	.644	.195	.308	.416
.3	.525	.638	.683	.164	.315	.423
.4	.440	.644	.655	.127	.293	.403
.5	.346	.642	.659	.084	.296	.407

Note. * 0 = no change, .5 = half the distance of the ceiling.

Discussion

Difference scores lose power for comparing the changes between two groups when variance decreases from pretest to posttest, due to either the presence of a floor/ceiling effect, or a skewed distribution combined with a mean change. ANCOVA, in contrast, is unaffected by either of these conditions, and retained power in spite of the decreased posttest variance. These findings are consistent with those of Stoolmiller and Bank (1995) who also found that difference scores lost power, relative to ANCOVA, when variance decreased from pretest to posttest. The second set of simulations showed that the effects of decreased posttest variance due to both floor/ceiling and skewness generalized to a continuous predictor. The regression based measure of change (partial correlation) was unaffected by the decrease in variance, while the correlation with difference scores exhibited a loss of power with decreases in posttest variance. The pattern of the relationship between the latent predictor and change variables from the structural equation model was dependent on the manner in which the relationship was assessed. Specifically, although the value of the correlation coefficient was affected by the decreased variance (i.e., the correlation decreased), the power associated with the covariance (using the associated maximum likelihood estimates of the covariance and standard error) was unaffected. These findings indicate that ANCOVA (or other regression based methods such as partial correlation) and structural equation models of change are to be preferred to difference scores when strong floor/ceiling effects are present at posttest or when a skewed distribution is changing in the direction of decreasing variance.

The explanation for the loss of power shown by ANOVA on difference scores when posttest variability decreases lies in the smaller change signal, relative to the pretest error variance. The ratio of the change signal to the error in the posttest will remain constant as the posttest variability decreases, but the absolute value of the change signal will decrease, relative to the size of the pretest variance. ANOVA on difference scores lacks power to detect this small signal because the pretest error variance and posttest error variance are pooled. The formula for the error variance of difference scores, $s^2_{difference} = s^2_{pre} + s^2_{post} - 2r_{pre,post}s_{pre}s_{post}$, contains equal weighting for the pretest and posttest error variances, where s^2 represents the sample variance, s represents the sample standard deviation and r represents the Pearson correlation coefficient. In contrast, regression based methods (ANCOVA or partial correlation) do not pool the pretest and posttest variances and are able to detect this small signal embedded in the smaller posttest variance. The methodological connection between difference scores and latent change variables (with two time points) in SEM (see Duncan et al., 1999) may explain why the correlation from the structural equation model of change is also affected by the decreasing variance.

Table 6

Average Gain Score Correlation, Average Partial Correlation, Average Correlation Between the Predictor and Change in the Structural Model, and Average Fit Indices for the Structural Model When Scores Approach a Ceiling

Reliability (of pretest) = .86							
Change	Gain Scores	Partial r	SEM r	GFI	AGFI	RMSEA	CFI
0	.160	.161	.196	.985	.966	.019	.998
.1	.157	.161	.194	.985	.966	.019	.998
.2	.149	.159	.188	.985	.965	.020	.998
.3	.143	.162	.182	.985	.966	.018	.998
.4	.128	.160	.162	.985	.966	.019	.998
.5	.110	.160	.139	.985	.966	.018	.998
Reliability (of pretest) = .61							
Change	Gain Scores	Partial r	SEM r	GFI	AGFI	RMSEA	CFI
0	.099	.106	.188	.985	.966	.019	.996
.1	.098	.109	.192	.985	.966	.018	.996
.2	.094	.110	.191	.985	.965	.020	.996
.3	.087	.108	.180	.985	.965	.020	.996
.4	.076	.106	.158	.985	.965	.020	.996
.5	.066	.105	.132	.985	.965	.019	.996

Note. GFI = Goodness of fit index; AGFI = Adjusted goodness of fit index; RMSEA = Root mean square error of approximation; CFI = Comparative fit index.

Table 7

Average Power Rates (Proportion Significant) for Gain Scores, Partial r and the SEM Path Measuring the Correlation With Change, When Scores Change Toward the Head of a Negatively Skewed Distribution

	Reliability (of pretest)					
	.86			.61		
Change	Gain Scores	Partial r	SEM r	Gain Scores	Partial r	SEM r
0	.644	.651	.689	.230	.282	.399
25	.565	.628	.650	.181	.299	.434
50	.491	.648	.674	.164	.319	.440
75	.404	.645	.666	.115	.311	.416
100	.323	.674	.727	.102	.287	.405

These findings extend the work of Bonate (2000) who explored the effect of increasing posttest variance and concluded it had little effect on the power of either difference scores or ANCOVA. We also find that increasing posttest variance does not differentially affect these methods (results not presented here). Bonate examined several non-normal (skewed) distributions and failed to detect differences between ANCOVA and difference scores. However, he did not explore conditions involving a mean shift, i.e., the means of both groups increasing (or decreasing) from pretest to posttest. Thus his findings are comparable to the zero mean change condition in the present study. It is important to emphasize that non-normality (skewness) only impairs the power of difference scores when there is a mean change (away from the tail) which results in decreased posttest variance, a condition analogous to a floor/ceiling effect.

The present investigation only examined the effect of decreasing variance from pretest to posttest on power, not on rates of Type I error. Neither of the conditions examined here, a ceiling effect and a skewed distribution changing in the direction of decreased posttest variance, have any differential effect on the Type I error rates of difference scores and ANCOVA (Jamieson & Cribbie, in preparation). In general, changes in posttest variance have minimal effects on Type I error rates (e.g., Bonate, 2000). However, under certain conditions, skewness and changes in variance cause substantial increases in Type I errors for difference scores (Jamieson & Cribbie, in preparation).

Table 8

Average Gain Score Correlation, Average Partial Correlation, Average Correlation Between the Predictor and Change in the Structural Model, and Average Fit Indices for the Structural Model When Scores Change Toward the Head of a Negatively Skewed Distribution

Reliability (of pretest) = .86							
Change	Gain Scores	Partial r	SEM r	GFI	AGFI	RMSEA	CFI
0	.159	.160	.194	.984	.964	.020	.998
25	.147	.157	.185	.985	.965	.020	.998
50	.138	.162	.176	.985	.965	.019	.998
75	.122	.160	.156	.985	.965	.020	.998
100	.111	.164	.140	.984	.965	.020	.998
Reliability (of pretest) = .61							
Change	Gain Scores	Partial r	SEM r	GFI	AGFI	RMSEA	CFI
0	.100	.106	.186	.985	.964	.020	.996
25	.093	.108	.184	.985	.964	.021	.996
50	.086	.110	.173	.985	.965	.019	.996
75	.073	.108	.151	.985	.965	.020	.996
100	.067	.106	.132	.985	.965	.019	.996

The present findings confirm the problems with difference scores when posttest variability decreases (Collins 1996b), and show that structural equation models of change do not share this problem. These findings are of practical value to researchers, since they show that regression-based measures of change (ANCOVA, partial correlation) and structural equation models of change are superior to difference scores when posttest variability decreases. Further, a logical extension of this research would be to explore multiple group SEM (e.g., Byrne, 2001), which would be a direct generalization of the ANCOVA model, and would be appropriate for cases when the predictor of change is discrete and latent measures of change are of interest (i.e., multiple predictors of change are available) or there are more than two time points over which change is to be measured.

References

Bentler, P. M. (1988). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bonate, P. L. (2000). *Analysis of pretest-posttest designs.* Boca Raton: Chapman & Hall.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural models*. New Park: Sage Publications.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. Mahwah, NJ: LEA.

Collins, L. M. (1996a). Is reliability obsolete? A commentary on 'Are simple gain scores obsolete?' *Applied Psychological Measurement, 20*, 289-292.

Collins, L. M. (1996b). Measurement of change in research on aging: Old and new issues from an individual growth perspective. In J. E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of the aging* (4th ed.). San Diego, CA: Academic Press.

Cribbie, R. A., & Jamieson, J. (2000). Structural equation models and the regression bias for measuring correlates of change. *Educational and Psychological Measurement, 60*, 893-907.

Cronbach, J. & Furby, L. (1970). How should we measure change - Or should we? *Psychological Bulletin., 105*, 68-80.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* Mahwah, NJ: LEA.

Jamieson, J. (1995). Measurement of change and the law of initial values: A computer simulation study. *Educational and Psychological Measurement, 55*, 38-46.

Jamieson, J. (1999) Dealing with baseline differences: Two principles and two dilemmas. *International Journal of Psychophysiology, 31*, 155-161.

Jöreskog, K. & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications.* Chicago: SPSS, Inc.

Llabre, M. M., Spitzer, S. S., Saab, P. G., Ironson, G. H., & Schneiderman, N. (1991). The reliability and specificity of delta versus residualized change as measures of cardiovascular reactivity to behavioral challenges. *Psychophysiology, 28*, 701-711.

Lord, F. E. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*, 304-305.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32*, 215-253.

Maris, E. (1998). Covariance adjustment versus gain scores revisited. *Psychological Methods, 3*, 309-327.

Miller, G. A. & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*, 40-48.

Raykov, T. (1994). Studying correlates and predictors of longitudinal change using structural equation modeling. *Applied Psychological Measurement, 18*, 63-77.

Raykov, T. (1997). Simultaneous study of individual and group patterns of latent longitudinal change using structural equation modeling. *Structural Equation Modeling, 4*, 212-236.

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

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

SAS Institute (1990a). *SAS/STAT user guide* (Version 6, 4th ed.). Cary, NC: Author.

SAS Institute (1990b). *SAS procedures guide* (Version 6, 3rd ed.). Cary, NC: Author.

Schafer, W. D. (1992). Analysis of pretest-posttest designs. *Measurement and Evaluation in Counselling and Development, 25*, 2-4.

Steyer, R., Eid, M., and Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research Online 2*(1), 21-33. Retrieved July 01, 2004, from http://www.dgps.de/fachgruppen/methoden/mpr-online/

Stoolmiller, M., & Bank, L. (1995). Autoregressive effects in structural equation models: We see some problems. In J. M.. Gottman (Ed.), *The analysis of change* (pp. 261-276). Mahwah, NJ: Erlbaum.

Wainer, H. (1991). Adjusting for differential base rates: Lord's Paradox again. *Psychological Bulletin, 109*, 147-151.

Willett, J. B. (1997). Measuring change: What individual growth buys you. In E. Amsel & K. A. Reninger (Eds.), *Change and development*, pp. 213-243. Mahwah, NJ: Lawrence Erlbaum Associates.

Williams, R. H. & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement, 20*, 59-69.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.