# Evaluating Performance Improvement through Repeated Measures: A Primer for Educators Considering Univariate and Multivariate Designs

Kim Nimon
University of North Texas

Cynthia Williams
Texas Christian University

**Abstract**

Repeated measures analysis is an important tool for educators committed to evaluating the performance of their students and courses. While evaluations can be performed using a series of t-tests, repeated measures provides practitioners and researchers a more sophisticated tool to analyze the impact of education over time or interventions that employ concurrent tests to measure a particular set of knowledge, skills, or attitudes. This paper provides educators with the information they need to choose between and interpret results based on the univariate and multivariate approach to repeated measures analyses. It also serves to explain the sphericity assumption and its impact on repeated measures designs.

Keywords: univariate, primer, sphericity assumption, repeated measures, evaluation, research design

## Introduction

In order to evaluate learning programs, relevant skills, knowledge, and attitudes from program participants are often measured multiple times (Kirkpatrick & Kirkpatrick, 2006). For example, participants may be measured on their ability to perform a particular skill: (a) before taking a course (i.e., pre), (b) immediately after completing a course (i.e., post and retro), and (c) one month after completing a course (i.e., follow-up). To determine if there is a statistical or practical difference between these measurements, a series of t-tests could be conducted (e.g., post-pre, after-post, follow-up-pre). However, the results from such a procedure would be difficult to collectively interpret as the process does not provide for a single omnibus test (R. Henson, personal communication, April 19, 2006). Additionally, the process inflates familywise Type I error rate. This means that the reported probability levels would actually overestimate the statistical significance of the mean differences (Hinkle, Wiersma, Jurs, 2003).

A more appropriate technique to analyze three or more measurements is the repeated measures design (Maxwell & Delaney, 2004). Repeated measures designs are also called within-subjects designs (Girden, 1992). In the case where the design contains a between-subjects factor in addition to a within-subjects factor, the design may be called a mixed-model, randomized block, or a split-plot design (Lamb, 2003). This paper presents a within-subjects repeated measures design with one within-subjects factor and no between-subjects factor (i.e., one-way within-subjects design). Readers interested in more advanced repeated measures designs are directed to Maxwell and Delaney (2004) and Stevens (2002).

Two approaches for implementing a one-way within-subjects design are discussed: (a) univariate, and (b) multivariate. Before presenting the two approaches, advantages and disadvantages of repeated measures are reviewed as well as the underlying statistical assumptions for the two techniques. The paper concludes by summarizing the differences between the univariate and multivariate approaches.

## Advantages and Disadvantages

*Advantages*

Maxwell and Delaney (2004) cited two advantages of within-subjects design: (a) sample size and (b) precision. In the case of a repeated measures design, each subject contributes $n$ scores, where $n$ equals the number of measurements. In the example previously depicted, $n$ equals 4. As a result of each subject contributing $n$ scores, the number of subjects needed to achieve a certain level of statistical power is often much lower in within-subjects designs than in between-subject designs where participants contribute only one score on the dependent variable (Maxwell & Delaney). Venter and Maxwell (1999) showed that in the case of a two-level design, the total number of subjects $N_W$ needed for the within-subjects design is related to $N_B$, the total number of subjects in the between-subject design, as follows:

$$N_W = N_B (1 - \rho)/2 \quad (1)$$

where $\rho$ is the population correlation between scores at the two levels of the within-subjects design. Table 1 further illustrates the sample size benefits of a one-way two-level repeated measures design.

<Insert Table 1 about here>

It is important to note that Venter and Maxwell's (1999) formula relies on compound symmetry and is therefore most applicable to the univariate approach to repeated measures.

However, in the case of a two-level design, the univariate and multivariate approaches are identical (Maxwell & Delaney, 2004). Therefore, the sample size benefits of a one-way two-level repeated measures design are identical for both repeated measures techniques (i.e., univariate and multivariate). A generalization of Venter and Maxwell's formula is presented in the univariate section of this paper. Considerations for determining sample size in a multivariate analysis are presented in the multivariate section.

In addition to requiring fewer subjects than between-subjects designs, repeated measures designs provide greater precision since subjects serve as their own control (Stevens, 2002). Because comparisons in the repeated-measures designs are made within-subjects, variability in individual differences between-subjects is removed from the error term (Maxwell & Delaney, 2004). Figure 1 illustrates this point. In the repeated measures design, the error term ($SS_{Res}$) does not include the variable among individuals ($SS_I$) as its counterpart ($SS_W$) does in the between-subjects design. As the variance among individuals is partitioned out of the error term, repeated measures designs are much more powerful than completely randomized designs (Stevens, 2002) and most likely result in a larger eta-squared (K. Roberts, personal communication, July 5, 2004).

<Insert Figure 1 about here>

*Disadvantages*

Tanguma (1999) identified three disadvantages of repeated measures design: (a) practice effects, (b) differential carryover effects, and (c) the potential for violations of statistical assumptions. Descriptions of the first two disadvantages and techniques for management are discussed. As conforming to the underlying statistical assumptions is a critical issue for all research designs (Hinkle, Wiersma, & Jurs, 2003), such issues are reserved to a subsequent section of the paper devoted to the subject.

*Practice effects*

Practice effects occur when subjects change systematically during the course of an experiment (Tanguma, 1999). Such changes may involve a positive or negative effect (Lewis, 1993).

In the case of education, a positive practice effect may indicate an improvement in subjects' knowledge, skills, or attitudes. However, in lieu of a learning program being responsible for the change, the improvement may be an artifact of the participants being retested using the same or similar instrumentation (Gall, Gall, & Borg, 2003). A technique to manage practice effects is to integrate a control group into the repeated measures design since the re-testing effect should manifest itself equally in the control and the experimental group (Campbell & Stanley, 1963).

Tanguma(1999) indicated that a negative practice effect may result from fatigue or boredom. He recommended that researchers lengthen the rest period between measurement occasions to manage fatigue and provide incentives as a technique to motivate participants throughout the course of the experiment.

Counterbalancing is also identified as a technique to manage practice effects (Lamb, 2003; Maxwell & Delaney, 2004; Tanguma, 1999; Wells, 1998). However, counterbalancing is most appropriate for designs where subjects are observed in different treatment conditions (Maxwell & Delaney) as counterbalancing is a way of ordering treatments so that each treatment is administered an equal number of times first, second, third, and so on, in particular sequences of conditions given to different subjects (Tanguma) . In the case of evaluating the effects of a learning program, participants are usually subjected to one treatment and then observed

longitudinally over time (Kirkpatrick & Kirkpatrick, 2006). Counterbalancing therefore would not be an appropriate technique to manage the practice order effects when measuring participants using the traditional occasions (e.g., pre, post, retro, and follow-up).

*Differential carryover effects*

An artifact of counterbalancing may be differential carryover effects. "Differential carryover effect occurs when the carryover effect of Treatment Condition 1 onto Treatment Condition 2 is different from the carryover effect of Treatment Condition 2 onto Treatment Condition 1" (Maxwell & Delaney, 2004, p. 556). Tanguma (1999) asserted that a possible solution to differential carryover effects is providing participants sufficient time between treatments so that the treatment condition dissipates completely from the subjects' system. Maxwell and Delany disagree and assert that a within-subjects design should be abandoned if differential carryover effect is a potential threat to validity. For the typical learning program evaluation, differential carryover effects is not an issue since implementing a counterbalanced design is not appropriate for reasons previously stated.

**Statistical Assumptions**

Stevens (2002) identified three assumptions for a single-group repeated measures analysis: (a) independence of observations, (b) multivariate normality, and (c) sphericity. Of the three assumptions, the first two apply to the multivariate approach while all three apply to the univariate approach.

*Independence of observations*

Violation of independence of observations can lead to increased Type I error rate (Hinkle, Wiersma, & Jurs, 2003). While this assumption is typically met through random selection (Gall, Gall, & Borg, 2003), learning programs are usually evaluated with intact groups. The interaction of the group may affect the scores of the members resulting in correlated observations (Lamb, 2003). Correlated observations can cause an overestimation of the true probability and is resolved by testing at a more stringent level of significance (Stevens, 2002).

*Multivariate normality*

The properties of ANOVA and MANOVA that make them robust to violations of multivariate normality carry over to repeated measures designs (Stevens, 2002). However, statistical tests of sphericity are not robust to the assumption of multivariate normality (Olejnik & Huberty, 1993). In the absence of multivariate normality, statistical tests of sphericity may indicate heterogeneity of variance between measurement occasions when they should fail to reject the null hypothesis (Minke, 1997). See Henson (1999) for techniques to assess multivariate normality.

*Sphericity*

*Testing for Sphericity.* Simply stated, the sphericity assumption is met when the variance at each measurement occasion is equal (K. Roberts, personal communication, July 5, 2004). Girden (1992) identified two techniques to test for sphericity: (a) examining variances of differences between all pairs of measurement occasions and (b) examining the matrix of orthonormal contrasts.

*Variances of Differences between Pairs of Measurement Occasions.* The variance of differences between two measurement occasions can be computed using the following formula (Girden, 1992):

$$\sigma^2{}_{A-B} = \sigma^2{}_A + \sigma^2{}_B - 2\sigma_{AB} \quad (2)$$

where $\sigma^2_A$ is the variance of a set of scores under measurement occasion A, $\sigma^2_B$ is the variance of a set of scores under measurement occasion B, and $\sigma^2_{AB}$ is the covariance of the two sets of scores. The more direct way of determining variance between two occasions is to compute the variance of the difference scores (Girden). Using either technique, sphericity is met if the variances between all pairs of measurement occasions are equal (Tanguma, 1999).

Using the variance-covariance information in Table 3 based on the heuristic data in Table 2, $\sigma^2_{A-B} = 79.817$, $\sigma^2_{A-C} = 233.635$, $\sigma^2_{A-D} = 91.273$, $\sigma^2_{B-C} = 163.636$, $\sigma^2_{B-D} = 111.272$, $\sigma^2_{C-D} = 59.818$. Table 4 illustrates that the same variances are computed when using difference scores. For the data set identified in Table 2, the sphericity assumption is not met.

<Insert Table 2 about here>
<Insert Table 3 about here>
<Insert Table 4 about here>

*Matrix of Orthonormal Contrasts*

Girden (1992) and Stevens (2002) asserted that sphericity is also said to exist if:

$$C^T \sum C = \sigma^2 I \quad (3)$$

where $C$ is a matrix of ($k$ - 1) orthogonal contrasts, $C^T$ is the transpose of $C$, $\sum$ is the variance-covariance matrix, and $I$ is an identity matrix. Multiplying the matrix of orthogonal contrasts identified in Table 5, its transpose (Table 6), and the variance-covariance matrix for the data in Table 2 (Table 3) results in the covariance matrix of transformed variables depicted in Table 7. For the dataset illustrated, the sphericity assumption is not met as the covariance matrix for the transformed variables does not have equal variances on the diagonal (Stevens, 2002).

<Insert Table 5 about here>
<Insert Table 6 about here>
<Insert Table 7 about here>

*Mauchly's Sphericity Test.* Maxwell and Delaney (2004) highlighted that while sphericity tests such as the techniques outlined by Girden (1992) indicate variance inequalities in the sample, the sphericity assumption is only violated if it holds in the population as well. The authors recognized that even if sample variances are unequal, such inequalities might simply reflect sampling error. Therefore, they recommended that Mauchly's sphericity test (i.e., Mauchly's *W*) be used to test the null hypothesis that the homogeneity condition holds in the population.

While Mauchly's *W* has limitations in behavioral science research (including the analysis of learning program outcomes) due to its sensitivity to multivariate normality (Stevens, 2002), it is presented here since the results of the test are automatically generated in software packages (e.g., SPSS 14.0 for Windows) that conduct repeated measures analyses. Furthermore, studies conducted by Huynh and Mandeville (as cited in Keselman, Rogan, Mendoza, & Breen, 1980) found that for short-tailed distributions, the test basically maintains the true rate of Type I error below the level of significance alpha.

Figure 2 depicts the results for the Mauchly's test for the dataset represented in Table 2. The results are interpreted the same way as Levene's test for homogeneity of variance in ANOVA. If the *p-calc* value generated is greater than or equal to the *p-crit* value defined by the researcher, then homogeneity of variance is assumed. Otherwise, the sphericity assumption is not met. In the example provided, Mauchly's test indicates that the heterogeneity of variance between measurement occasions is statistically significant at the .05 alpha level ($p = .018$).

<Insert Figure 2 about here>

*Managing Violations to Sphericity.* If the sphericity assumption is not met, the *F* ratio generated by the univariate repeated measures analysis is positively biased, rejecting falsely too often (Maxwell & Delaney, 2004). For example, if the alpha level is set at .05 and the sphericity assumption is not, univariate repeated measures analyses may falsely reject the null hypothesis 10% or 15% of the time (Stevens, 2002). To adjust for the positive bias, the degrees of freedom for the repeated measures *F* test may be corrected using one of three adjustments: (a) Greenhouse-Geisser, (b) Huynh-Feldt, and (c) Lower-bound. However, it is important to note that while the adjusted tests provide better control for Type I error rate, they are only approximate (Maxwell & Delaney).

The Greenhouse-Geiser formula shown below results in a parameter ($\hat{\varepsilon}$) that identifies the extent to which the covariance matrix deviates from sphericity (Stevens, 2002):

$$\frac{a^2(\overline{E}_{jj} - \overline{E})^2}{(a-1)((\sum\sum E^2_{jk}) - (2a\sum \overline{E}^2_{j.}) + (a^2\overline{E}^2_{..}))} \quad (4)$$

where $E_{jk}$ is the element in row $j$ and column $k$ of the sample covariance matrix, $\overline{E}_{jj}$ is the mean of variances along the diagonal in the sample covariance matrix, $\overline{E}_{j.}$ is the mean of all entries in $j^{th}$ row of the sample covariance matrix, $\overline{E}_{..}$ is the mean of all entries in the sample covariance matrix, and $a$ is the number of measurement occasions. The resulting parameter is used to correct the degrees of freedom for the measurement occasion and error term. For the dataset depicted in Table 2, $\hat{\varepsilon}$ is .610. Applying the $\hat{\varepsilon}$ to the unadjusted degrees of freedom for the measurement occasion ($(a - 1) = 3$) and the error term ($(n - 1) * (a - 1) = 33$) results in corrected degrees of freedom of 1.820 and 20.115, respectively.

The Huynh-Feld formula results in a parameter ($\tilde{\varepsilon}$) that identifies the extent to which the covariance matrix deviates from sphericity (Stevens, 2002):

$$\frac{n(a-1)\hat{\varepsilon} - 2}{(a-1)(n-1-(a-1)\hat{\varepsilon})} \quad (5)$$

where $n$ is the number of subjects, $a$ is the number of measurement occasions, and $\hat{\varepsilon}$ is the Greenhouse-Geisser adjustment. The resulting parameter is used to correct the degrees of freedom for the measurement occasion and error term. For the dataset depicted in Table 2, $\tilde{\varepsilon}$ is .725. Applying the $\tilde{\varepsilon}$ to the degrees of freedom for the measurement occasion and the error term results in corrected degrees of freedom of 2.175 and 23.920, respectively.

The lower-bound adjustment simply sets the degrees of freedom for the measurement occasion to one and the degrees of freedom for the error term to ($n$ - 1). The lower-bound adjustment suggests that no matter how badly the homogeneity assumption is violated, the largest possible critical *F* value needed requires one and ($n$ – 1) degrees of freedom (Maxwell & Delaney, 2004).

Figure 3 illustrates the associated effect on the *p*-value for each of the three adjustments. Of the three techniques, the Greenhouse-Geiser formula provides a moderate correction, the

Huynh-Feld is the least conservative, and the lower-bound adjustment is the most conservative. The Greenhouse-Geiser formula tends to underestimate $\varepsilon$, while the Huyn-Feld adjustment tends to overestimate $\varepsilon$ (Stevens, 2002). Therefore, Stevens recommended that in lieu of using any of these three adjustments directly that researchers use the average of the Greenhouse-Geisser and Huyn-Feld adjustments in order to correct the degrees of freedom for the repeated measures *F* test. Alternatively, he indicated that researchers choose the Greenhouse-Geisser test to be *somewhat conservative.*

<Insert Figure 3 about here>

**Univariate**

In presenting the univariate approach to repeated measures, the following tasks are considered: (a) calculating sample size, (b) conducting the omnibus test, (c) computing effect size, (d) analyzing contrasts, and (e) reporting results. The topics are presented in approximate procedural order.

*Calculating Sample Size*

Although Cohen's classical text (1988) on power analysis provides power tables for a variety of situations, it does not provide tables for repeated measures. However, formulas for determining the appropriate sample size for a single group repeated measures design can be derived after first determining the sample size needed for a between-subjects design (Stevens, 2002). The following is one such formula (Maxwell & Delaney, 2004):

$$N_W = N_B\,(1-\rho)/a \tag{6}$$

where $N_W$ equals the sample size for the within-subjects design, $N_B$ is the sample size for the between-subjects design, $\rho$ is the average correlation for the subjects' responses to all measurement occasions, and $a$ is the number of measurement occasions. It is important to note that the formula relies heavily on sphericity. In cases where the sphericity assumption is not met, researchers are directed to Elashoff (as cited in Maxwell & Delaney, 2004).

*Conducting Omnibus Test*

The univariate repeated measures omnibus test for a single group compares an *F-calc* to an *F-crit* similar to a between-subjects ANOVA. However, the difference between the two approaches relates to variation among individuals: First, the denominator of the *F-calc* (error term) excludes the variation among individuals. Second, the degrees of freedom for the error term excludes the degrees of freedom associated with individuals. Table 8 outlines the formulas for computing the repeated measures *F-calc*. Table 9 depicts their use based on the data identified in Table 2 assuming that the sphericity assumption has been met.

<Insert Table 8 about here>

<Insert Table 9 about here>

The univariate technique for conducting a repeated measures omnibus test for a single group can also be conducted using a statistical software package. Figure 4 identifies the SPSS code to conduct a repeated measure test for the data identified in Table 2. Figure 5 relates relevant output to an ANOVA summary table consistent with the information provided in Table 9.

<Insert Figure 4 about here>

<Insert Figure 5 about here>

*Computing Effect Size*

In addition to determining the statistical significance of a univariate repeated measures design, it is also important to analyze the practical significance of the test (Henson, in press). This can be accomplished by computing omega squared ( $\omega^2$ ). The formula for $\omega^2$ in one-way within-subjects designs based on the univariate approach is as follows (Maxwell & Delaney, 2004):

$$\omega^2 = \frac{(k-1)(MS_{occasions} - MS_{error})}{SS_{total} + MS_{individuals}} \quad (7)$$

where *k* equals the number of measurement occaions, *MS* denotes mean square, and *SS* denotes sums of squares. Applying these formulas to the data in Table 2 results in an $\omega^2$ of .0377, indicating that the measurement occasion accounted for 3.77% of the variance in the dependent variable.

*Analyzing Contrasts*

In addition or in lieu of conducting a univariate repeated measures omnibus test (Oljenik & Huberty, 1993), researchers may want to analyze specific means differences or conduct trend analyses. In either case, this is accomplished by testing contrasts. The univariate formula for testing contrasts is as follows (Maxwell & Delaney, 2004):

$$F_{calc} = n\overline{D}^2 / MS_{error} \quad (8)$$

where *n* equals the number of subjects, *D* is the transformed variable resulting from applying the contrasts to the original data, and $MS_{error}$ is the pooled average error term generated by the omnibus test. As the univariate formula employs a pooled error term, it relies heavily on the sphericity assumption. If the assumption is not met, $MS_{error}$ should be replaced with an individual error term. Testing contrasts with a separate variance estimate approach is consistent with multivariate analyses. Therefore, its formula is outlined in the multivariate section.

To illustrate the process of conducting a trend analysis, a contrast matrix is identified in Table 10. Applying the contrast matrix elements to the data in Table 2 results in a set of transformed variables identified in Table 11. Applying the formulas to the transformed variables indicates that the linear and quadratic trends are not statistically significant ($F_{linear}(1,11) = 3.19$; $p > .05$ and $F_{quadratic}(1,11) = .20$, $p > .05$). However, the cubic trend is statistically significant ($F_{cubic}(1,11) = 5.69$; $p < .05$).

<Insert Table 10 about here>

<Insert Table 11 about here>

Statistical software packages also report the results of polynomial trends as a byproduct of conducting a repeated measures analysis. Figure 6 outlines the relevant trend analysis output generated by SPSS for the data in Table 2. However, SPSS employs a separate variance estimate approach in lieu of the pooled error term. Therefore, the *F* values generated by SPSS are different than the hand calculations previously noted ($F_{linear}(1,11) = 2.475$; $p = .144$; $F_{quadratic}(1,11) = .219$; $p = .649$; $F_{cubic}(1,11) = 7.066$; $p = .022$).

<Insert Figure 6 about here>

*Reporting Results*

In addition to reporting the ANOVA summary table (as depicted in Table 9 and Figure 5), researchers need to report on results of a priori tests, null hypothesis tests, effect size calculations, and post-hoc analyses (Henson, in press; Ojenick & Huberty, 1993). The following provides an example write-up of the results of the tests conducted for the data in Table 2. The data obtained from the four points of measurement lacked sphericity (Mauchly's $W = .034$; $p = .018$). Therefore, the Greenhouse-Geisser adjustment was employed in analyzing the repeated

measures ($\hat{\varepsilon}$ = .610). We fail to reject the null hypothesis that the amount of perceived knowledge measured at four different points of time relative to a learning intervention are equal ($F$(1.829, 20.115) = 3.027, $p$=.064). As indicated by the univariate $\omega^2$ (Maxwell & Delaney, 2004), occasion accounted for 3.77% of the variance in perceived knowledge. Trend analysis indicated that the cubic trend was statistically significant ($F$(1,11) = 7.066, $p$ = .022).

**Multivariate**

In presenting the multivariate approach to repeated measures, the following tasks are considered: (a) calculating sample size, (b) conducting the omnibus test, (c) computing effect size, (d) analyzing contrasts, and (e) reporting results. The topics are presented in approximate procedural order.

*Calculating Sample Size*

Maxwell and Delaney (2004) outlined sample size tables for conducing repeated measures analyses using the multivariate approach (pp. 640-643). The authors indicated that the values were obtained by using a noncentrality parameter value of:

$$\delta^2 = nd^2 / 2(1 - \rho_{\min}) \quad (9)$$

where $n$ equals the sample size, $d$ is the expected effect, and $\rho_{\min}$ is the minimum correlation between measurement occasions. They further noted four patterns to the tables: First, the required number of subjects generally increases as the number of levels increases. Second, the number of subjects increases as the level of desired power increases. Third, as $d$ increases, the number of subjects needed decreases. Fourth, as $\rho_{\min}$ increases, the number of subjects decreases as higher correlations are indicative of greater consistency in subjects' scores across measurement occasions making effects easier to detect.

When considering the sample size requirements for a multivariate test compared to a univariate test, the multivariate approach is less powerful in the presence of sphericity (Stevens, 2002). Maxwell and Delaney (2004) also noted that all other things being equal, the multivariate approach loses power when compared to the univariate approach, as the number of subjects ($n$) decreases. They further asserted that the multivariate approach may be mathematically impossible when $n$ is less than the number of levels ($k$) + 10. However, in cases where $n$ is greater than $k$ + 10 and there is a large violation of sphericity ($\varepsilon$ < 0.7), the multivariate procedure is more powerful (Field, n. d.).

*Conducting Omnibus Test*

Hotelling's $T^2$ is consistently used (e.g., Girden, 1992; Stevens, 2002; Tanguma, 1999) as the multivariate statistic to analyze repeated measures. It is important to note that the multivariate analysis is not performed on the original scores but on the differences between adjacent measurements (Tanguma). Table 11 identifies the latent variables constructed for the data in Table 2.

As the following formulas show, Hotelling's $T^2$ (formula 11) is analogous to the $t$ statistic (formula 10) for dependent samples:

$$t^2 = \frac{\bar{d}^2}{s_d^2 / n} \quad (10)$$

where $d$ is the mean difference between two dependent samples and $s_d^2$ is the variance of difference scores, and $n$ is the number of subjects.

$$T^2 = n y'_d S_d^{-1} y_d \quad (11)$$

where $n$ is the number of subjects, y'$_d$ is the row vector of mean differences on the ($k$ - 1) difference variables, $S_d$ is the matrix of variances and covariances on the ($k$ – 1) difference variables (Stevens, 2002).

As depicted in Figure 7, $T^2$ for the data in Table 2 is 8.21. Applying the following formula that converts $T^2$ to an $F$ statistic results in an $F_{calc}$ of 2.24:

$$F = [(n - k + 1) / ((n - 1)*(k - 1))] T^2 \quad (12)$$

where $n$ equals the number of subjects and $k$ equals the number of measurement occasions. The resultant $F_{calc}$ is insufficient to reject the null hypothesis at the .05 alpha level with 3 ($k$ – 1) and 9 ($n$ – $k$ – 1) degrees of freedom.

<Insert Figure 7 about here>

The multivariate technique for conducting the repeated measures omnibus test for a single group can also be conducted using a statistical software package. Figure 8 identifies the SPSS code to conduct a multivariate repeated measures test for the data identified in Table 2. Figure 9 identifies relevant SPSS output. Note that the Hotelling trace coefficient (.74670) depicted in Figure 9 is a derivative of the $T^2$ previously computed, where:

$$\text{Hotelling trace coefficient} = T^2 /(n - 1) \quad (13)$$

where $n$ equals the number of subjects. Also note that the $F$ statistic identified (2.24) is the same as the $F_{calc}$ previously computed. The multivariate tests (Pillais, Hotellings, and Wilks) conducted also provide identical $F$ statistics and $p$-values. Chen (2004) indicated that while the tests usually provide similar results, Wilks' output should be chosen in the event the results are different.

<Insert Figure 8 about here>

<Insert Figure 9 about here>

*Analyzing Contrasts*

While the omnibus multivariate repeated measures test is performed on latent variables, the multivariate approach to testing contrasts is performed on the original scores. The process mirrors the univariate approach. The only exception is that the error term in the multivariate approach is an individual error term such that the multivariate formula for testing contrasts is as follows (Maxwell & Delaney, 2004):

$$F_{calc} = n\overline{D}^2 / S^2_D \quad (14)$$

where $n$ equals the number of subjects, $D$ is the transformed variable resulting from applying the contrasts to the original data, $S^2_D$ is the variance for the vector of transformed variables.

Applying the multivariate formula to the transformed variables identified in Table 11 indicates that the linear and quadratic trends are not statistically significant ($F_{linear}(1,11) = 2.475$; $p > .05$ and $F_{quadratic}(1,11) = .219$, $p > .05$). However, the cubic trend is statistically significant ($F_{cubic}(1,11) = 7.066$; $p < .05$).

Statistical software packages also report the results of polynomial trends as a byproduct of conducting a multivariate repeated measures analysis. Figure 10 outlines relevant trend analysis output generated by SPSS MANOVA command for the data in Table 2. While the univariate analyses are based on $t$-values (i.e., $t_{linear} = 1.572$, $t_{quadtratic} = .467$, $t_{cubic} = -2.658$), the $p$-values generated (i.e., $p_{linear} = .144$, $p_{quadtratic} = .649$; $p_{cubic} = .022$) are the same as those generated from the univariate $F$ tests resulting from the General Linear Model (GLM) command (see Figure 6). This illustrates that SPSS employs a multivariate approach (i.e., an individual error term) when testing contrasts as a consequence of the GLM or the MANOVA command.

<Insert Figure 10 about here>

*Computing Effect Size*

In addition to determining the statistical significance of a multivariate repeated measures design, it is also important to analyze the practical significance of the test (Henson, in press). This can be accomplished by computing omega squared ( $\omega^2$ ). The formula for $\omega^2$ in one-way within-subjects designs based on the multivariate approach is as follows (Maxwell & Delaney, 2004):

$$\omega^2 = 1 - \frac{n\Lambda}{df_{error} + \Lambda} \tag{15}$$

where *n* equals the number of subjects, *df* denotes degrees of freedom, and $\Lambda$ equals the Wilks' lambda value. Applying this formula to the multivariate results in Figure 9 results in an $\omega^2$ of .282, indicating that the measurement occasion accounted for 28.2% of the variance in the composite dependent variable. It is important to note that the multivariate omega squared is approximately seven times larger than the univariate omega squared for the same data. While this may appear to be an advantage of the multivariate approach, total variance is conceptualized differently between the two approaches. In particular, variation attributable to systematic individual differences is excluded from the total variance in the multivariate conceptualization (Maxwell & Delaney, 2004). Maxwell and Delaney asserted that since variability due to subjects should be included in the conceptualization of total variance, the univariate version of omega squared is preferred.

*Reporting Results*

The following provides an example write-up of the results of the multivariate approach to testing the repeated measures for the data in Table 2. Using Wilks' lambda criteria, we fail to reject the null hypothesis that the composite amount of perceived knowledge measured at four different points of time relative to a learning intervention are equal ($F(3,9) = 2.240$, $p$=.153). As indicated by the univariate $\omega^2$ (Maxwell & Delaney, 2004), occasion accounted for 3.77% of the variance in perceived knowledge. Trend analysis indicated that the cubic trend was statistically significant ($F(1,11) = 7.066$, $p = .022$).

**Summary**

In considering the differences between the multivariate and univariate approaches to repeated measures analyses, Maxwell and Delaney noted four issues: (a) statistical assumptions, (b) tests of contrasts, (c) Type I error rate, and (d) Type II error rate (power). After summarizing the differences between the univariate and multivariate considerations for each of these subjects, this paper concludes by presenting guidelines to use when considering the two approaches.

*Statistical assumptions*

The distinction between the statistical assumptions required for the two approaches is sphericity. While the sphericity assumption is not applicable to the multivariate approach, the univariate approach assumes sphericity. In particular, the univariate approach to conducting omnibus tests, contrast tests, and sample size calculations requires sphericity.

*Tests of Contrasts*

Testing contrasts in the multivariate approach employs individual error terms, while the univariate approach employs a pooled error term. Therefore, the univariate approach to testing contrasts can provide misleading results when the sphericity assumption is violated.

*Type I Error Rate*

Type I error rate can be two to three times higher than the nominal value in the univariate approach when sphericity is violated. While $\varepsilon$ adjustments provide better control, they are not exact. The multivariate approach produces exact Type I error rates assuming that its statistical assumptions have been met (Maxwell & Delaney, 2004).

*Type II Error Rate*

Under the condition of sphericity, univariate tests provide better power than the multivariate approach. When sphericity is not met, neither test is uniformly more powerful than the other. However, as the degree of violation of sphericity increases, the power for the multivariate test increases.

*Guidelines*

Faced with the differences between the univariate and multivariate approaches, Field (n.d.) identified the following rules of thumb for choosing between univariate and multivariate approach to repeated measures analyses: (a) The multivariate approach is preferred when there is a large violation of sphericity ($\varepsilon < 0.7$) and when $n$ is greater than ($k + 10$). (b) The univariate approach is preferred when sphericity holds ($\varepsilon > 0.7$) or when the sample size is small.

Stevens (2002) provided a different guideline for considering a repeated measures approach. He indicated that if researchers can meet Maxwell and Delaney's (2004) rule of thumb relating sample size to number of levels ($n > k + 10$) that they conduct *both* the adjusted univariate and multivariate test and discern any differences in treatment effects. He further recommended that researchers following this advice set the experimentwise level of significance for each test to half of the overall desired alpha level.

.

References

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi- Experimental Designs for Research*. Boston, MA: Houghton Mifflin Company.

Chen, Y. H. (2004). Repeated measures analysis. *Singapore Medical Journal, 45*(8), 354-369.

Field, A. (n. d.). A bluffer's guide to … sphericity. *BPM-MSC Newsletter 6*(1). Retrieved April 26, 2006 from http://www.sussex.ac.uk/Users/andyf/research/articles/sphericity.pdf

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston, MA: A B Longman.

Girden, E. R. (1992). *ANOVA: Repeated measures.* Newbury Park, CA: Sage.

Henson, R. (1999). Multivariate normality: What is it and how can it be assessed? In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 233 – 250). Stanford, CT: JAI Press.

Henson, R. (in press). Effect size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist.*

Hinkle, D. E., Wiersma, W., & Jurs. S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin Company.

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*, 479-481.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluation Training Programs: The Four Levels* (3rd ed.). San Francisco, Berrett-Koehler Publishers, Inc.

Lamb, G. D. (2003, February). *Understanding within versus between ANOVA designs: Benefits and requirements of repeated measures.* Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, Texas.

Lewis, C. (1993). Analyzing means from repeated measures design. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Statistical issues* (pp. 73-94). Hillsdale, NJ: Erlbaum.

Maxwell, S. E., & Delany, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associated, Publishers.

Minke, A.(1997, January). *Conducted repeated measures analyses: Experimental design considerations.* Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, Texas.

Olejnik, S., & Huberty, C. J. (1993, April). *Preliminary statistical tests.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, Georgia.

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

Tanguma, J. (1999). Repeated measures: A primer. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 233 – 250). Stanford, CT: JAI Press.

Venter, A., & Maxwell, S. E. (1999). Maximizing power in randomized designs when *N* is small. In R. H. Hoyle (ed.), *Statistical strategies for small sample research* (pp. 31-58). Thousand Oaks, CA: Sage.

Wells, R. D. (1998, November). *Conducting repeated measures analyses using regression: The general linear model lives*. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, Louisiana.

Table 1.
*Sample Size Required to Detect a Medium Difference between Two Means (power = .80)*

| $\rho$ | Between-Subjects | Within-Subjects |
|---|---|---|
| 0.0 | 128 | 64 |
| 0.3 | 128 | 45 |
| 0.5 | 128 | 32 |
| 0.7 | 128 | 20 |

*Source:* Maxwell & Delaney (2004, p. 562).

Table 2.
*Heuristic Dataset*

| Subject | A | B | C | D | Mean |
|---|---|---|---|---|---|
| 1 | 96 | 108 | 122 | 110 | 109 |
| 2 | 117 | 103 | 133 | 127 | 120 |
| 3 | 107 | 96 | 107 | 106 | 104 |
| 4 | 85 | 84 | 99 | 92 | 90 |
| 5 | 125 | 118 | 116 | 125 | 121 |
| 6 | 107 | 110 | 91 | 96 | 101 |
| 7 | 128 | 129 | 128 | 123 | 127 |
| 8 | 84 | 90 | 113 | 101 | 97 |
| 9 | 104 | 84 | 88 | 100 | 94 |
| 10 | 100 | 96 | 105 | 103 | 101 |
| 11 | 114 | 105 | 112 | 105 | 109 |
| 12 | 117 | 113 | 130 | 132 | 123 |
| Total | 1284 | 1236 | 1344 | 1320 | |

Note: The grand mean ($M_{grand}$) of the 48 scores is 108.

Table 3.
*Variance-Covariance Matrix for Data in Table 2*

| | A | B | C | D |
|---|---|---|---|---|
| A | 200.545 | 154.364 | 97.455 | 143.636 |
| B | 154.364 | 188.000 | 121.182 | 127.364 |
| C | 97.455 | 121.182 | 218.000 | 168.091 |
| D | 143.636 | 127.364 | 168.091 | 178.000 |

Table 4.
*Variance of Difference Scores for Data in Table 2*

| Subject | A-B | A-C | A-D | B-C | B-D | C-D |
|---|---|---|---|---|---|---|
| 1 | -12.000 | -26.000 | -14.000 | -14.000 | -2.000 | 12.000 |
| 2 | 14.000 | -16.000 | -10.000 | -30.000 | -24.000 | 6.000 |
| 3 | 11.000 | 0.000 | 1.000 | -11.000 | -10.000 | 1.000 |
| 4 | 1.000 | -14.000 | -7.000 | -15.000 | -8.000 | 7.000 |
| 5 | 7.000 | 9.000 | 0.000 | 2.000 | -7.000 | -9.000 |
| 6 | -3.000 | 16.000 | 11.000 | 19.000 | 14.000 | -5.000 |
| 7 | -1.000 | 0.000 | 5.000 | 1.000 | 6.000 | 5.000 |
| 8 | -6.000 | -29.000 | -17.000 | -23.000 | -11.000 | 12.000 |
| 9 | 20.000 | 16.000 | 4.000 | -4.000 | -16.000 | -12.000 |
| 10 | 4.000 | -5.000 | -3.000 | -9.000 | -7.000 | 2.000 |
| 11 | 9.000 | 2.000 | 9.000 | -7.000 | 0.000 | 7.000 |
| 12 | 4.000 | -13.000 | -15.000 | -17.000 | -19.000 | -2.000 |
| Variance | 79.818 | 223.636 | 91.273 | 163.636 | 111.273 | 59.818 |

Table 5.
*Matrix of Orthonormal Contrasts for Data in Table 2*

| Occasion | C1 | C2 | C3 |
|---|---|---|---|
| A | .707 | .408 | .289 |
| B | -.707 | .408 | .289 |
| C | .000 | -.816 | .289 |
| D | .000 | .000 | -.866 |

Table 6.
*Transpose of Matrix identified in Table 5*

| Occasion | A | B | C | D |
|---|---|---|---|---|
| C1 | .707 | -.707 | .000 | .000 |
| C2 | .408 | .408 | -.816 | .000 |
| C3 | .289 | .289 | .289 | -.866 |

Table 7.
*Covariance Matrix of Transformed Variables for Data in Table 2*

| | T1 | T2 | T3 |
|---|---|---|---|
| T1 | 39.898 | 17.308 | -12.248 |
| T2 | 17.307 | 115.649 | 28.060 |
| T3 | -12.248 | 28.060 | 26.675 |

Table 8.
*Formulas for Conducting the Repeated Measures Omnibus Test*

| Source | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| Occasions | $\sum(T^2/n)-(G^2/N)$ | *k-1* | $SS_{occasions}/df_{occasions}$ | $MS_{occasions}/MS_{error}$ |
| Individuals | $\sum k(M_{subject}-M_{grand})^2$ | *n-1* | $SS_{individuals}/df_{individuals}$ | |
| Error | $SS_{total} - SS_{individuals} - SS_{occasions}$ | *(k-1)(n-1)* | $SS_{error}/df_{error}$ | |
| Total | $\sum X^2-(G^2/N)$ | *N-1* | | |

Note: $T$ = sum of the test scores for each particular test, $G$ = sum of all the scores; $\sum X^2$ = sum of all squared scores; $N$ = number of scores in the entire experiment; $M_{subject}$ = mean of each individual's scores; $M_{grand}$ = grand mean of all scores; $n$ = number of individuals; $k$ = number of occasions

Table 9.
*Repeated Measures ANOVA Summary Table and Related Computations for Data in Table 2*

| Source | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| Individuals | $4*[(109-108)^2 + (120-108)^2 + (104-108)^2 + (90-108)^2 + (121-108)^2 + (101-108)^2 + (127-108)^2 + (97-108)^2 + (94-108)^2 + (101-108)^2 + (109-108)^2 + (123-108)^2] = 6624$ | 12-1= 11 | 602.18 | |
| Occasions | $[(1284^2/12) + (1236^2/12)+ (1344^2/12) + (1320^2/12)] – (5{,}184^2/48) = 552$ | 4-1= 3 | 184.00 | 3.03 |
| Error | $9182 – 6624 – 552 = 2006$ | (4-1)*(12-1)= 33 | 60.79 | |
| Total | $569054 – (5184^2/48) = 9182$ | 48-1 = 47 | | |

Note: *F-crit* (3,33) ~= 2.84; therefore, the null hypothesis is rejected at the .05 alpha level.

Table 10.
*Matrix of Orthonormal Contrasts to Analyze Polynomial Trends for Data in Table 2*

| | Contrasts | | |
|---|---|---|---|
| Measurement | Linear | Quadratic | Cubic |
| A | -0.671 | 0.500 | -0.224 |
| B | -0.224 | -0.500 | 0.671 |
| C | 0.224 | -0.500 | -0.671 |
| D | 0.671 | 0.500 | -.224 |

Table 11.
*Transformed Variables and Latent Variables based on Data in Table 2*

| Subject | Linear | Quadratic | Cubic | A-B | B-C | C-D |
|---|---|---|---|---|---|---|
| 1 | 12.52 | -12.00 | -6.26 | -12 | -14 | 12 |
| 2 | 13.42 | 4.00 | -17.89 | 14 | -30 | 6 |
| 3 | 1.79 | 5.00 | -7.60 | 11 | -11 | 1 |
| 4 | 8.05 | -3.00 | -8.50 | 1 | -15 | 7 |
| 5 | -0.45 | 8.00 | 1.34 | 7 | 2 | -9 |
| 6 | -11.63 | 1.00 | 10.29 | -3 | 19 | -5 |
| 7 | -3.58 | -3.00 | -0.45 | -1 | 1 | 5 |
| 8 | 16.55 | -9.00 | -11.63 | -6 | -23 | 12 |
| 9 | -1.79 | 16.00 | -3.58 | 20 | -4 | -12 |
| 10 | 4.02 | 1.00 | -5.37 | 4 | -9 | 2 |
| 11 | -4.47 | 1.00 | -6.71 | 9 | -7 | 7 |
| 12 | 13.86 | 3.00 | -8.05 | 4 | -17 | -2 |
| Mean | 4.02 | 1.00 | -5.37 | 4 | -9 | 2 |
| Variance | 78.56 | 54.91 | 48.92 | | | |

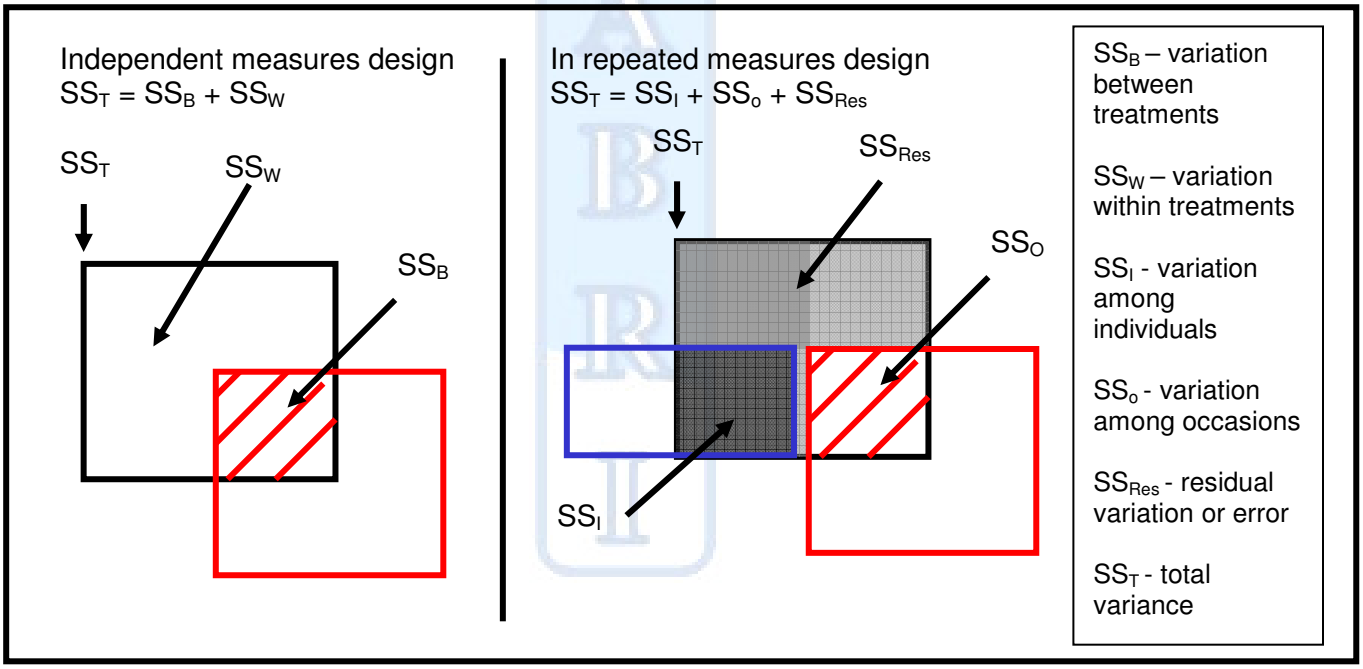


*Figure 1.* Comparison of Sum of Squares Partitioning between Designs (K. Roberts, personal communication, July 5, 2004).

**Mauchly's Test of Sphericity(b)**

Measure: MEASURE_1

| Within Subjects Effect | Mauchly's W | Approx. Chi-Square | df | Sig. | Epsilon(a) | | |
|---|---|---|---|---|---|---|---|
| | | | | | Greenhouse-Geisser | Huynh-Feldt | Lower-bound |
| occassion | .243 | 13.768 | 5 | .018 | .610 | .725 | .333 |

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b Design: Intercept
Within Subjects Design: occassion

*Figure 2.* Results of Mauchly's Test of Sphericity for Data in Table 2.

**Tests of Within-Subjects Effects**

Measure: MEASURE_1

| Source | | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| occassion | Sphericity Assumed | 552.000 | 3 | 184.000 | 3.027 | .043 |
| | Greenhouse-Geisser | 552.000 | 1.829 | 301.865 | 3.027 | .075 |
| | Huynh-Feldt | 552.000 | 2.175 | 253.846 | 3.027 | .064 |
| | Lower-bound | 552.000 | 1.000 | 552.000 | 3.027 | .110 |
| Error(occassion) | Sphericity Assumed | 2006.000 | 33 | 60.788 | | |
| | Greenhouse-Geisser | 2006.000 | 20.115 | 99.727 | | |
| | Huynh-Feldt | 2006.000 | 23.920 | 83.863 | | |
| | Lower-bound | 2006.000 | 11.000 | 182.364 | | |

*Figure 3.* Univariate F Test Results for Data in Table 2.

```
GLM
 A B C D
  /WSFACTOR = occassion 4 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = occassion .
```

*Figure 4.* SPSS Code to Conduct Repeated Measures Analyses for Data in Table 2.

**Tests of Within-Subjects Effects**

Measure: MEASURE_1

| Source | | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| occassion | Sphericity Assumed | 552.000 | 3 | 184.000 | 3.027 | .043 |
| | Greenhouse-Geisser | 552.000 | 1.829 | 301.865 | 3.027 | .075 |
| | Huynh-Feldt | 552.000 | 2.175 | 253.846 | 3.027 | .064 |
| | Lower-bound | 552.000 | 1.000 | 552.000 | 3.027 | .110 |
| Error(occassion) | Sphericity Assumed | 2006.000 | 33 | 60.788 | | |
| | Greenhouse-Geisser | 2006.000 | 20.115 | 99.727 | | |
| | Huynh-Feldt | 2006.000 | 23.920 | 83.863 | | |
| | Lower-bound | 2006.000 | 11.000 | 182.364 | | |

**Tests of Between-Subjects Effects**

Measure: MEASURE_1
Transformed Variable: Average

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Intercept | 559872.000 | 1 | 559872.000 | 929.739 | .000 |
| Error | 6624.000 | 11 | 602.182 | | |

| Source | *SS* | *df* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Individuals | 6624 | 11 | 602.18 | | |
| Occasions | 552 | 3 | 184.00 | 3.03 | .043 |
| Error | 2006 | 33 | 60.79 | | |
| Total | 9182 | 47 | | | |

*Figure 5.* Relevant Univariate SPSS Output and ANOVA Summary Table for Data in Table 2.

**Tests of Within-Subjects Contrasts**

Measure: MEASURE_1

| Source | factor1 | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| factor1 | Linear | 194.400 | 1 | 194.400 | 2.475 | .144 |
| | Quadratic | 12.000 | 1 | 12.000 | .219 | .649 |
| | Cubic | 345.600 | 1 | 345.600 | 7.066 | .022 |
| Error(factor1) | Linear | 864.000 | 11 | 78.545 | | |
| | Quadratic | 604.000 | 11 | 54.909 | | |
| | Cubic | 538.000 | 11 | 48.909 | | |

*Figure 6.* Trend Analysis SPSS Output (GLM command) for Data in Table 2.

$$T^2 = 12 \begin{bmatrix} 4 & -9 & 2 \end{bmatrix} \begin{pmatrix} 79.82 & -9.91 & -40.00 \\ -9.91 & 163.64 & -56.09 \\ -40.00 & -56.09 & 59.82 \end{pmatrix} \begin{pmatrix} 4 \\ -9 \\ 2 \end{pmatrix}$$

$$= 8.21$$

*Figure 7.* $T^2$ Computations for Latent Variables in Table 12.

```
MANOVA A B C D
 /WSFACTORS=Measure(4)
 /CONTRAST(Measure)=POLYNOMIAL
 /PRINT= SIGNIF(AVERF) TRANSFORM.
```

*Figure 8.* SPSS Code to Conduct Multivariate Repeated Measures Analyses for Data in Table 2.

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
EFFECT .. MEASURE
 Multivariate Tests of Significance (S = 1, M = 1/2, N = 3 1/2)
Test Name        Value    Exact F Hypoth. DF   Error DF  Sig. of F
Pillais           .42749    2.24010       3.00       9.00       .153
 Hotellings       .74670    2.24010       3.00       9.00       .153
 Wilks            .57251    2.24010       3.00       9.00       .153
 Roys             .42749
 Note.. F statistics are exact.
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

*Figure 9.* Multivariate Repeated Measures SPSS Output for Data in Table 2.

```
Estimates for T2
 --- Individual univariate .9500 confidence intervals
MEASURE
  Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper
        1   4.02492236    2.55841    1.57321     .14397   -1.60610    9.65594
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Estimates for T3
 --- Individual univariate .9500 confidence intervals
MEASURE
  Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper
        1   1.00000000    2.13910     .46749     .64928   -3.70813    5.70813
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Estimates for T4
 --- Individual univariate .9500 confidence intervals
MEASURE
  Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper
        1   -5.3665631    2.01885   -2.65823     .02226   -9.81002    -.92310
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

*Figure 10.* Trend Analysis SPSS Output (MANOVA command) for Data in Table 2.