**Glossary of Terms**

**B**: B is the non-standardized regression weight.

**Bivariate analysis**: Bivariate analysis involves the examination of the relationship between two variables. Classic examples of bivariate analyses would be the Pearson correlation between two variables and the independent samples t-test (which examines the impact of an independent variable on a dependent variable).

** Canonical analysis**: Whereas standard correlation examines the relationship between two variables, canonical correlation analysis examines the relationship (correlation) between two sets of variables.

** Canonical variate**: The canonical variate results from canonical correlation analysis. The canonical variate is a function in the form of a linear combination of two or more variables.

**Cluster analysis**: Cluster analysis statistically sorts cases into groups (the dependent variable) based on similarity or dissimilarity on a set multiple predictor (independent) variables with others in the group.

**Confounder**: A confounder is a type of control variable that researchers expect to interfere with the clear determination of the relationship between the independent and dependent variable(s).

**Control variable (also referred to as covariate)**: Control variables (covariates) are independent variables that are included in the model. One would interpret an analysis with control variables (covariates) by examining the statistical effect of the independent variable on the dependent variable when controlling for the influence of covariates also included in the model.

**Cross-loaded items**: Cross-loaded items are those items loading similarly on multiple factors.

**Dependent variable (DV)**: A dependent variable depends on the independent variable. The manipulation of the IV causes a presumed effect on the DV. Other terms for dependent variable used in the literature are criterion and outcome variables.

**Discriminant analysis**: Discriminant analysis (also referred to as discriminant function analysis) examines a pattern of multiple predictor (independent) variables that best discriminates two or more levels of a dependent variable or combinations of more than one dependent variable. The analysis is a test used to verify or discover combinations of independent variables that predict group membership.

**Effect size**: The effect size is a measure of the strength of association between variables. It allows researchers to a method of quantifying differences between two groups.

**Eigenvalue**: The eigenvalue provides a measure of overall importance of the canonical variate. In general, larger eigenvalues denote more important canonical variates.

**Factor Analysis**: Factor analysis reduces a large number of variables into a smaller set of items that are highly correlated with each other.

**Homogeneity of variance**: Homogeneity of variance is an assumption of parametric statistical tests. If met, two or more groups (represented by levels of the variable) are equal.

**Homogeneity of variance-covariance matrices**: Just as homogeneity of variance is an assumption for parametric statistical tests, the equivalent in the multivariate case is the homogeneity of variance-covariance matrices.

**Independent variable (IV)**: In experimental research, the researcher manipulates the independent variable. The IV causes an expected effect in another variable. Other terms for independent variable used in the literature are predictor variable and grouping variable.
**Linear regression**: Linear regression examines the effect of multiple independent (predictor) variables on a single dependent (criterion) variable. In linear regression, the outcome (dependent or criterion variable) is an interval or ratio.

**Linearity**: Linearity means there is a straight-line relationship between pairs of variables. If relationships between variables are nonlinear, Type II error can increase. If tests show that curvilinear relationships exist, then the simplest solution is to add curvilinear terms (squaring the value of the independent variable) as predictors to capture that variance associated with the curvilinear relationship. This also serves to increase power.

**Logistic regression**: In logistic regression, the dependent or criterion variable is typically categorical. When the outcome has two categories or levels, it is binary logistic regression.

**Model fit**: For logistic regression, a fit index (a common one is the Hosmer-Lemeshow test) assesses overall fit of the model to the data. A test that is not statistically significant indicates a good fit between the data and the model.

**Multicollinearity**: Multicollinearity occurs when two independent variables are highly correlated.

**Multivariate analysis**: Multivariate analysis examines the relationships among multiple independent and dependent variables. Examples of these include factor analysis and structural equation modeling.

**Multivariate analysis of variance**: Multivariate analysis of variance (MANOVA) tests for mean differences among multiple dependent variables (two or more) when there is one or more independent variables.

**Multivariate outliers**: Multivariate outliers can result in violations of normality and linearity. Outliers are scores that are very different from the majority of the scores. In large samples, it is often possible to remove outliers to correct any nonlinear and non-normal situations.

**Multivariate normality**: There is no commonly available or agreed-upon test for multivariate normality (Stevens, 2002). Studies have shown that even in small sample sizes (20 to 40), with skewness as extreme as ± 2.0 and kurtosis as extreme as + 6.0, the Type I error rate is within .02 of the rate expected for normally distributed data. Large negative kurtosis (- 2.0) can increase Type II error. In general, larger sample sizes (greater than 100) are more robust to violations of normality.

**Odds ratio (OR)**: The odds ratio is an estimate of the change in the odds of membership in a group for a one-unit change in the predictor variable. This is how one most frequently sees the results of logistic regression models presented.

**R²**: R² is a measure of effect size for multiple regression and is the proportion of variance explained by the predictors (IV) in the outcome variable (DV).

**ΔR²**: The change in R² is the change in overall variance explained in the outcome when a new independent variable is added to the model. Change in R² can be particularly useful in describing the importance of adding a new independent variable (or a group of variables) to the model.

**Scree plot**: The scree plot contains the factors graphed as a function of eigenvalues. One looks at the plot and makes a subjective call to determine the number of factors to retain in factor analysis based on where the curve abruptly changes. Just before this point would be the number of factors to retain.

**Semi-partial correlation**: The semi-partial correlation squared is the proportion of variance in the DV explained solely by a particular predictor (when controlling for the effects of all other predictors in the model).
Squared canonical correlation: The proportion of variance in the weighted set of learning styles scores that is associated with weighted intelligence scores.

Standardized beta (β): Standardized beta results when the regression coefficient B is standardized to be on a scale of 0 to +/- 1.0. This allows for easier interpretability of relative importance of predictors in the model.

Standardized discrimination coefficients: Standardized discrimination coefficients allow for the interpretation of the contribution of each measure to the canonical variate.

Structural equation modeling (SEM): Structural equation modeling evaluates the goodness of fit of a hypothesized model to the data. These models vary in complexity and combine the analysis methods and advantages of factor analysis, known as confirmatory factor analysis (CFA), and regression. Typically, researchers use SEM to examine more complex models than can be analyzed with factor analysis or regression separately.

Structure loadings: Structure loadings are standardized correlations of items to their own factor. They have a magnitude that ranges from 0.0 to 1.0. Items that are more important have higher loadings.

Univariate analysis: Univariate analysis involves analysis of a single variable. Examples of this would be descriptive analyses of a variable, such as age.

Unstandardized discriminant coefficients: Unstandardized discriminant coefficients allow the researcher to create mathematical equations that describe each of the canonical variates (two in this case) that emerge.

Wilks’ lambda (Λ): Wilks’ lambda (Λ) provides statistical evidence for an overall multivariate effect of type of college major on the combined set of learning style scores. Λ is similar to the overall F-test in the analysis of variance procedure. Similar to the overall F-test, however, one cannot determine the nature of the differences when Λ is statistically significant. These require follow-up tests. Values for Λ vary between 0 and 1.0.