SUPPLEMENTAL CHAPTER:
MULTIVARIATE EFFECT SIZE ESTIMATION

This chapter supplements the presentations in Kline (2004, chaps. 4, 6-7) about effect size estimation in univariate designs.[1] A basic familiarity with the concepts presented in the source just mentioned and with multivariate methods of data analysis in general is assumed. See Stevens (2002) or Tabachnick and Fidell (2001) for an introduction to multivariate statistical techniques.

OVERVIEW

Recall that the analysis of variance (ANOVA) is just a special case of multiple regression (e.g., Keppel & Zedeck, 1989) and that both techniques are special cases of canonical correlation, which estimates the associations between multiple predictor variables and multiple outcome variables controlling for intercorrelations. Canonical correlation is the broadest technique in the general linear model, which also includes techniques such as the multivariate analysis of variance (MANOVA) and discriminant function analysis (DFA; also called predictive discriminant analysis; e.g., Hand & Taylor, 1987; Huberty, 1994; Silva & Stam, 1995). Relations between the techniques just mentioned are emphasized in the discussion that follows.

MULTIVARIATE EFFECT SIZE ESTIMATION IN TWO-GROUP DESIGNS

Two different effect size indexes for contrasts between two independent samples across multiple continuous outcome variables are introduced below. One is a multivariate extension of the standardized mean difference Hedges's *g* and the other is a multivariate version of the variance-accounted-for effect size estimated eta-squared, $\hat{\eta}^2$ (also called the correlation ratio). In order to understand these multivariate effect size indexes, one needs to know something about MANOVA. A brief, nontechnical overview follows.

In two-group designs, computer programs for MANOVA construct the weighted linear combination of the outcome variables that best separates the groups taking account of variable intercorrelation. This best combination is known as the *discriminant function* (DF) or *canonical variate*. Other descriptive terms include *composite* or *principal component*. One way to gauge the relative contribution of the individual outcome variables to group discrimination is to inspect their simple correlations with the DF in the total data set, which are called *structure coefficients*. (In exploratory factor analysis, standardized factor loadings are structure coefficients.) If a structure coefficient is about zero, that variable does not contribute very much to group discrimination; the opposite conclusion is drawn as the absolute value of a structure coefficient approaches 1.0. Structure coefficients tend to be more statistically stable than other measures of relative variable importance, such as standardized DF coefficients (weights), which are analogous to standardized regression coefficients (beta weights).

One measure of the discriminating power of the whole DF is Wilks's lambda (Λ), which equals

---

[1]This work supplements:

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association, esp. chaps. 4, 6–7.

Please reference this work as:

Kline, R. B. (2004). *Supplemental chapter: Mmultivariate effect size estimation*. Retrieved [date, e.g., June 20, 2004] from http://www.apa.org/books/resoources/kline

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{A}|} \tag{1}$$

where the numerator is the determinant of the pooled within-groups sum of squares cross-products (SSCP) matrix **W** and the denominator is the determinant of the matrix total of **W** and **A**, the between-groups SSCP matrix for the dichotomous *A* factor (i.e., the group contrast). A determinant is analogous to the variance of a matrix, which makes Λ a ratio of unexplained variability over total variability. The corresponding ratio in ANOVA is SS_W/SS_T where the numerator is the pooled within-groups sum of squares and the denominator is the total sum of squares for the whole design. As the magnitude of the multivariate group contrast gets larger, the value of Λ approaches zero. Wilks's Λ can be converted to *F* for a statistical test of the multivariate group contrast that assumes homogeneity of the within-groups variance-covariance matrices; see Stevens (2002) or Tabachnick and Fidell (2001) for suggestions about how to evaluate this assumption and possible corrective measures.

The *Mahalanobis generalized distance*, D^2_{M}, is a multivariate statistic that expresses in a squared metric the distance between the group *centroids*—the vectors of univariate means—relative to the pooled-within groups variance-covariance matrix. The square root of D^2_{M} is thus a type of multivariate standardized mean difference that can be seen as a multivariate extension of the univariate standardized mean difference Hedges's *g*. A general formula is

$$D^2_{\mathrm{M}} = \mathbf{g}'\mathbf{R}^{-1}\mathbf{g} \tag{2}$$

where **g′** and **g** are, respectively, row and column matrices (vectors) of *g* for the individual outcome variables and $\mathbf{R}^{-1}$ is the inverse of the pooled within-groups correlation matrix. If there are just two outcome variables, Equation 2 is reduced to

$$D^2_{\mathrm{M}} = \left(\frac{1}{1-r_p^2}\right)(g_1^2 + g_2^2 - 2\,r_p\,g_1\,g_2) \tag{3}$$

where and g_1 and g_2 are the univariate standardized mean differences for the individual outcome variables and r_p is the pooled within-groups correlation. The latter is just the partial correlation between the two outcome variables controlling for group membership coded as 0 or 1 (or any two different numbers). Negative values of r_p are associated with higher values of D^2_{M}. This is because the term "$2\,r_P\,g_1\,g_2$" in Equation 3 is a negative number when $r_p < 0$ and is a positive number when $r_p > 0$.

If there are more than two dependent variables, it is more convenient to calculate D^2_{M} as

$$D^2_{\mathrm{M}} = \frac{df_W\,(1-\Lambda)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{\Lambda} \tag{4}$$

where Λ is from the two-group MANOVA, n_1 and n_2 are the group sizes, and df_W is the usual ANOVA total within-groups degrees of freedom (i.e., $n_1 + n_2 - 2 = N - 2$). Because Λ reflects the discriminating power of the whole set of outcome variables, the value of the square root of D^2_{M} can be quite a bit higher than the absolute value of *g* for any individual outcome variable.

The *Wilks's generalized correlation ratio* is computed directly from Wilks's Λ for the multivariate group contrast. It is a multivariate extension of the correlation ratio $\hat{\eta}^2$ for the individual outcome variables. The formula for generalized correlation ratio is:

$$\text{mult. } \hat{\eta}^2 = 1 - \Lambda \tag{5}$$

(3)

Because Λ is a multivariate proportion of unexplained variance (see Equation 1), one minus Λ is a multivariate proportion of explained variance in the total data set. In two-group designs, mult. $\hat{\eta}^2$ is the proportion of variance in the DF explained by group membership. It also equals the squared *canonical correlation* between the dichotomous factor and the continuous DF. In other words, the square root of mult. $\hat{\eta}^2$ is basically a multivariate point-biserial correlation when there are only two groups.

For two reasons, the value of mult. $\hat{\eta}^2$ can be substantially higher than that of $\hat{\eta}^2$ for any of the individual outcome variables. First, the statistic Λ reflects the discriminating power of all the outcome variables considered together. Second, group membership explains only some proportion of the variance in the DF—which is estimated as mult. $\hat{\eta}^2$—and the DF in turn explains only some proportion of the variance in the set of outcome variables. The latter is calculated as the sum of the squared structure coefficients divided by the number of outcome variables.

The product of mult. $\hat{\eta}^2$ and the average squared structure coefficient is known as *redundancy*, and it refers to how much of the actual variance in one set of variables is explained by the other set (Stewart & Love, 1968), in this case group membership and the original set of outcome variables, not the DF based on them. Redundancy in a comparative study can be calculated more directly as the average of the $\hat{\eta}^2$ values across the individual outcome variables. In two-group designs, redundancy is the average proportion of variance in the set of outcome variables explained by group membership through the single DF. For the same data, redundancy may be substantially lower than mult. $\hat{\eta}^2$. This is because the two statistics have different reference points: the proportion of variance shared between group membership and the DF (mult. $\hat{\eta}^2$) versus the proportion of variance extracted in the set of outcome variables by the dichotomous factor that represents group membership (redundancy).

Table 1 presents raw scores and descriptive statistics for two groups and two outcome variables, *v*1 and *v*2. The pooled within-groups correlation is $r_p = -.28$, the univariate standardized mean differences are $g_1 = .80$ and $g_2 = 1.20$, and the values of the univariate correlation ratios are $\hat{\eta}_1 = .41$ and $\hat{\eta}_2 = .56$. (Readers should verify these results using the formulas presented in Kline, 2004, chap. 4.) Results of the two-group MANOVA for these data are:

$$\Lambda = .5298,\ F\,(2, 7) = 3.11, \text{ and } p = .108$$

The structure coefficients for v_1 and v_2 in the total sample are, respectively, .60 and .81. These results indicate that *v*2 contributes somewhat more to group discrimination than *v*1. The value of the Mahalanobis generalized distance between the group centroids is calculated for these data as:

$$\begin{aligned} D^2_{\mathrm{M}} &= 1/[1 - (-.28^2)]\ [.80^2 + 1.20^2 - 2\ (-.28)\ (.80)\ (1.20)] \\ &= [8\ (1 - .5298)\ (1/5 + 1/5)]/.5298 = 2.84 \end{aligned}$$

Thus, the multivariate standardized mean difference in an unsquared metric equals $2.84^{1/2}$, or 1.69.

Insert Table 1 about here

Wilks's generalized correlation ratio for data in Table 1 equals

$$\text{mult. } \hat{\eta}^2 = 1 - .5298 = .4702$$

so we can say the percentage of variance in the DF explained by group membership is about 47.0% and their bivariate correlation is approximately $.47^{1/2} = .69$. Redundancy for this example can be computed either as the average of the $\hat{\eta}^2$ values for *v*1 and *v*2 or as the product of mult. $\hat{\eta}^2$ and the average squared structure coefficient:

$$\text{Redundancy} = (.41^2 + .56^2)/2 = .47\ (.60^2 + .81^2)/2 = .24$$

Thus, the multivariate percentage of variance in the set of outcome variables *v*1 and *v*2 explained by the dichotomous factor of group membership is about 24%. Readers are encouraged to reproduce the above results with a computer program for MANOVA or canonical correlation.

The multivariate effect size indexes just described are subject to the same general limitations as their univariate counterparts (see Kline, 2004, chap. 4); both also assume homogeneity of the within-groups variance-covariance matrices. There is little point in using either multivariate effect size index if intercorrelations among the outcome variables are all about zero. In this case, each variable measures an independent facet of the group difference, so it makes no sense to analyze a linear combination of them. See Huberty (1994) and Olejnik and Algina (2000) for additional computational examples for multivariate two-group contrasts. The distributions of D_M^2 and mult. $\hat{\eta}^2$ are complex, and the author is not aware of a computer program that constructs confidence intervals based on either multivariate effect size index.

TWO-GROUP DESIGN EXAMPLE

School psychologists in the United States may be obliged by federal or state laws to incorporate parental observations when children are tested for possible placement in special education. Kline, Lachar, and Boersma (1993) evaluated whether parental ratings on an objective measure of child cognitive and adjustment status, the Personality Inventory for Children (PIC; Wirt, Lachar, Klinedinst, & Seat, 1984), could discriminate between regular and special education students. The PIC has 12 substantive scales normed by child age and gender, where higher standard scores ($\mu = 50.00$, $\sigma = 10.00$) indicate greater problems. For this example, the total data set of 771 cases (regular education $n_1 = 199$, special education $n_2 = 572$) was randomly split into an analysis sample ($n_a = 335$; M_a age = 10.2 years, $s_a = 2.3$ years; 72% boys, 29% girls) and a replication sample ($n_b = 336$; M_b age = 10.1 years, $s_b = 2.4$ years; 69% boys, 31% girls). Descriptive statistics (means, standard deviations, correlations) for all samples and groups just mentioned are reported in Appendix A.

Reported in Table 2 are group means and standard deviations on each PIC scale in the analysis sample. Also reported in the table are values of r_{pb}, which estimate the bivariate association between each PIC scale and group membership (regular vs. special education). (Recall that r_{pb} is a special case of $\hat{\eta}^2$ for a dichotomous factor.) As expected, the highest bivariate correlations (range = .52-72) are found on scales that measure cognitive or learning problems, Achievement, Development, and Intellectual Screening. The results of a two-group MANOVA for the analysis sample only are:

$$\Lambda = .4566,\ F\ (12, 322) = 31.93, \text{ and } p < .001$$

The *F* and *p* values are not very interesting due to the relatively large sample size (i.e., the group contrast is expected to be statistically significant).

Insert Table 2 about here

The value of the generalized correlation ratio in the analysis sample is

$$\text{mult. } \hat{\eta}^2 = 1 - .4566 = .5434$$

and its square root equals $.5434^{1/2}$, or .7371. Based on these results, we can say that the multivariate percentage of explained variance is about 54.3% and that the correlation between group membership and the best linear composite of PIC scales (i.e., the DF) is about .74. The structure coefficients, which are also reported in Table 2, indicate that the three PIC cognitive scales contribute the most to group discrimination. This is not surprising, but other scales that measure poor social skills, disorganized behavior, or adjustment problems also have relatively high structure coefficients. In other words, the multivariate separation of the two groups, regular education and special education, does not appear to be based on a unitary construct. The average structure coefficient in Table 2 is .31. Redundancy can be calculated as the product of this average and 1 − .4566, which equals .31 (.5434), or .17. That is, the comparison of the regular and special education groups in the analysis sample explains about 17% of the variance in the set of PIC scales.

A case-level effect size analysis for this example is considered next. (Various options for case-level comparisons of two groups are discussed in Kline, 2004, chap. 4). The MANOVA results just described assume homogeneity of the within-groups variance-covariance matrices. Although the within-groups correlation matrices are actually quite similar (Appendix A), the special education group is more variable on some scales than the regular education group (see Table 2). Consequently, case classification in a DFA was based on the separate within-groups variance-covariance matrices. Cases in the analysis sample were classified first, and then the cases in the replication sample were classified using the functions from the analysis sample. The prior probabilities in both analyses were specified as .85 and .15 for, respectively, the regular education and special education groups, which are realistic base rates.

The classification results are summarized in Table 3. The overall hit rate in the analysis sample is about 84%, only somewhat lower than the overall hit rate of about 80% in the replication sample. These observed cross-group hit rates represent approximately a 70 to 75% reduction in the error rate expected under chance classification. Across both samples, the hit rate for the regular education cases is at least 90%. Classification hit rates are somewhat lower for the special education cases, about 75% in both samples. Most classification errors across both samples (109/123 in total) are false positives, which means that special educations cases were predicted to belong to the regular education group. Overall, these results suggest that the parental ratings of children in regular versus special education are reasonably distinct at both the group and case levels.

Insert Table 3 about here

In the second research example considered later, the special education sample is divided into three subsamples, learning disabled, emotionally impaired, and cognitively impaired, and these three groups are compared with the regular education students in a four-group comparison with a classification analysis. Before this second research example is considered, we must deal with prerequisite concepts about multivariate effect size estimation when there are three or more groups.

MULTIVARIATE EFFECT SIZE ESTIMATION IN BETWEEN-SUBJECTS ONE-WAY DESIGNS

The basic ideas of multivariate effect size estimation in two-group designs generalize to designs with three or more groups and a single fixed factor. (See Kline, 2004, chap. 6, for a review of contrasts in one-way designs.) The extension of Hedges's *g* for multivariate contrasts in such designs is the square root of the Mahalanobis generalized, distance defined as follows:

$$D^2_{\mathrm{M}_{\hat{\psi}}} = \mathbf{g}' \mathbf{R}^{-1} \mathbf{g} \qquad (6)$$

where $\mathbf{g}'$ and $\mathbf{g}$ are respectively row and column vectors of *g* for the individual outcome variables and $\mathbf{R}^{-1}$ is the inverse of the pooled-within groups correlation matrix for *all* groups, not just the ones involved in the contrast. It's probably easier to derive the generalized distance for a contrast as

$$D^2_{M_{\hat{\psi}}} = \frac{df_W \; (1 - \Lambda_{\hat{\psi}}) \left(\sum_{i=1}^{a} \frac{c_i^2}{n_i} \right)}{\Lambda_{\hat{\psi}}} \tag{7}$$

where a is the number of groups (i.e., levels of factor A); df_W is the ANOVA total within-groups degrees of freedom calculated as $N - a$; the terms c_i^2 and n_i are, respectively, the squared contrast weight and size of the ith group; and $\Lambda_{\hat{\psi}}$ is Wilks's lambda for the contrast, *not* for the omnibus effect. Equation 7 assumes a standard set of contrast weights (i.e., the sum of their absolute value is 2.0).

The statistic $\Lambda_{\hat{\psi}}$ is associated with the weighted combination of the outcome variables (the DF) that best separates the groups compared in the contrast. The equation is

$$\Lambda_{\hat{\psi}} = \frac{|\mathbf{W}|}{|\mathbf{W} + \hat{\boldsymbol{\psi}}|} \tag{8}$$

where the numerator is the determinant of the pooled within-groups SSCP matrix $\mathbf{W}$ and the denominator is the determinant of the matrix total of $\mathbf{W}$ and $\hat{\boldsymbol{\psi}}$, the contrast SSCP matrix. Because the denominator of $\Lambda_{\hat{\psi}}$ excludes all sources of between-groups variability except for that due to the contrast, the statistic $\Lambda_{\hat{\psi}}$ is a proportion of unexplained variance after all noncontrast effects have been removed from the total variance. The analogous ratio in one-way ANOVA is:

$$\frac{SS_W}{SS_W + SS_{\hat{\psi}}} = \frac{SS_W}{SS_T - SS_{\text{non-}\hat{\psi}}} \tag{9}$$

where $SS_{\text{non-}\hat{\psi}}$ is the sum of squares for all noncontrast sources of between-groups variability apart from that due to the contrast of interest.

The form of Wilks's generalized correlation ratio for a multivariate contrast in a design with at least three groups is:

$$\text{mult. } \hat{\eta}^2_{\hat{\psi}} = 1 - \Lambda_{\hat{\psi}} \tag{10}$$

Because the denominator of $\Lambda_{\hat{\psi}}$ represents error variance and the effect of just the contrast, mult. $\hat{\eta}^2_{\hat{\psi}}$ is actually a multivariate version of partial $\hat{\eta}^2_{\hat{\psi}}$, the partial correlation ratio for a univariate contrast. This implies that values of mult. $\hat{\eta}^2_{\hat{\psi}}$ are not generally additive over a set of contrasts, orthogonal or otherwise. The square root of mult. $\hat{\eta}^2_{\hat{\psi}}$ is the correlation between the contrast and its DF controlling for all noncontrast sources of between-groups variability.

The generalized correlation ratio for the omnibus effect—the simultaneous comparison of all levels of factor A (i.e., all groups are compared at once)—is calculated as

$$\text{mult. } \hat{\eta}^2_A = 1 - \Lambda_A \tag{11}$$

where Λ_A is Wilks's lambda for the omnibus comparison (i.e., from the a-group MANOVA). The equation for Λ_A is

$$\Lambda_A = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{A}|} \tag{12}$$

where $\mathbf{A}$ is the SSCP matrix for all the groups. (The analogous ratio in one-way ANOVA is SS_W/SS_T.) Because the denominator of Λ_A represents error variance plus all sources of between-groups variability, the quantity $1 - \Lambda_A$ is a multivariate proportion of total variance explained by variability among all levels of factor A.

For the same data, the value of mult. $\hat{\eta}^2_A$ can be substantially higher than that of $\hat{\eta}^2_A$ for any individual outcome variable. Huberty (1994) described an alternative multivariate variance-accounted-for effect size index referred to here as estimated tau-squared ($\hat{\tau}^2$) that may be more directly comparable with $\hat{\eta}^2_A$ for the individual outcome variables. Its equation is

$$\hat{\tau}^2 = 1 - \Lambda_A^{1/r} \tag{13}$$

where r is the *smaller* of the number of groups minus one (i.e., $df_A = a - 1$) or the number of outcome variables. If there are only two groups, $r = 1$ regardless of the number of variables and

$$\Lambda_A^{1/1} = \Lambda_A \text{ and } \hat{\tau}^2 = \text{mult. } \hat{\eta}^2_A$$

but if there are three or more groups in a multivariate design, then $r > 1$ and

$$\Lambda_A^{1/r} > \Lambda_A \text{ and } \hat{\tau}^2 < \text{mult. } \hat{\eta}^2_A$$

Some statistical programs, such as the General Linear Model (GLM) module in SPSS, actually print $\hat{\tau}^2$ instead of mult. $\hat{\eta}^2_A$ for omnibus multivariate comparisons. The value of $\hat{\tau}^2$ is one minus the geometric mean of the complements of all r squared canonical correlations. A geometric mean is the qth root of the product of q statistics. This point is elaborated below.

The variable r in Equation 13 refers to the number of *characteristic roots* in a one-way MANOVA. This is the maximum number of *pairs* of discriminant functions that could be constructed by the computer in the analysis of the omnibus effect. One member of each pair is a weighted combination of the outcome variables, and the other member is a weighted combination of the $a - 1$ dichotomous codes that represent the levels of factor A (i.e., group membership). The weights for each pair of composites are selected by the computer to maximize their Pearson correlation, which is the canonical correlation. The canonical correlations typically get smaller as additional pairs of functions are extracted. This happens because each new pair of functions is constructed within the residual variance unexplained by the previous pairs. It is also true that the r composites of the outcome variables are all pairwise uncorrelated. In other words, each weighted combination of the outcome variables represents an independent dimension along which the groups may vary. The proportion of the variance in each dimension explained by factor A is the square of its canonical correlation. The set of r squared canonical correlations is thus another way to describe the magnitude of the association between factor A and the outcome variables in a squared metric. Please note, though, that squared canonical correlations are not generally additive across pairs of functions.

Each of the r weighted combinations of the outcome variables typically extracts only some proportion of the variance in the individual measures. This proportion for each DF equals the average squared structure coefficient across all the outcome variables. The product of the squared canonical correlation and the average squared structure coefficient for each DF is redundancy, the average proportion of variance in the set of outcome variables explained by factor A through that DF. The *total redundancy* is the sum of the redundancies across the r pairs of discriminant functions. Total redundancy in a comparative study also equals the average of the $\hat{\eta}^2_A$ values for the individual outcome variables. Total redundancy is thus even another kind of variance-accounted-for effect size that can be computed for the omnibus effect.

Because it is easy to get lost among all the various multivariate measures of association just described, some recommendations for choosing among them are offered below:

1. A small number of contrasts can often be analyzed instead of the omnibus effect. This simplifies things because only

one DF is analyzed for each contrast. That is, the analysis is reduced to a series of two-group multivariate comparisons.

2. If the omnibus effect is analyzed but there is no interest in examining group differences on individual dimensions, then mult. $\hat{\eta}^2_A$ or $\hat{\tau}^2$ may be preferred as measures of association. Of the two, the latter is a more conservative choice and more directly comparable with univariate effect sizes.
3. The canonical correlations or redundancy may be of greater interest when it is important to evaluate the dimensionality of group differences (e.g., Lambert, Wildt, & Durand, 1988). That $\hat{\tau}^2$ is an average related to the squared canonical correlations would make this statistic of interest, too.

Another point warrants brief mention: Some computer programs allow rotation of the discriminant functions. This preserves the value of mult. $\hat{\eta}^2_A$ but may force the individual structure coefficients to head toward either 0 or 1.0 in absolute value. The same thing is routinely done in exploratory factor analysis for the same reason: to enhance the interpretation of linear combinations of observed variables. See Thomas (1992) for a discussion about the interpretation of discriminant functions.

Huberty (1994), Olejnik and Algina (2000), Schaffer and Gillo (1974), and Tatsuoka (1993) describe additional kinds of multivariate measures of associations for one-way or factorial designs. These include multivariate versions of estimated omega-squared ($\hat{\omega}^2$) and other bias-adjusted estimates of the population proportion of explained variance in addition to indexes based on test statistics other than Wilks's lambda, such as the Hotelling-Lawley trace or the Bartlett-Pallai trace. All three test statistics just mentioned and a fourth called Roy's greatest root can be converted to exact or approximate *F* ratios for statistical tests of the multivariate omnibus effect, but results of these tests do not always agree when there are more than two groups. Very briefly, tests based on the Bartlett-Pallai trace tend to be more powerful when the separation of the groups is distributed over at least the first two discriminant functions, and tests based on Roy's greatest root tend to be more powerful when the groups differ mainly on the first DF (e.g., Olson, 1976). The most widely used multivariate test statistic for one-way MANOVA is Wilks's lambda, which seems to offer a reasonable compromise in terms of statistical power. The choice between multivariate effect size indexes based on Wilks's lambda versus other test statistics should be based on considerations similar to those just mentioned.

Table 4 presents raw scores and descriptive statistics for three groups and two outcome variables, *v*1 and *v*2. The pooled within-groups correlation between *v*1 and *v*2 equals −.30. Demonstrated next is the computation of multivariate effect size indexes for the pairwise comparison $\hat{\psi}_1$ defined by the weights (1, 0, −1). The omnibus effect is analyzed in a research example considered later.

Insert Table 4 about here

The values of the contrast $\hat{\psi}_1$ for each individual outcome variable for the data in Table 4 are 1.00 for *v*1 and 3.80 for *v*2, and the pooled within-groups variances (i.e., MS_W) are 5.50 for *v*1 and 7.23 for *v*2. Values of the univariate test statistics, standardized mean differences, and partial correlations for this contrast are as follows:

	*v*1	*v*2
$t_{\hat{\psi}_1}$	.67	2.23
$g_{\hat{\psi}_1}$	.43	.53
partial $\hat{\eta}_{\hat{\psi}_1}$	.19	.40

The value of Wilks's lambda for the multivariate contrast $\hat{\psi}_1$ calculated by a computer program for one-way MANOVA is $\Lambda_{\hat{\psi}_1} =$ .6313. From this result we can derive

$$D_{M_{\hat{\psi}_1}} = \{12\ (1 - .6313)\ [1^2/5 + 0^2/5 + (-1^2)/5]/.6313\}^{1/2} = 2.80^{1/2}$$
$$= 1.67$$

which says that the distance between the centroid of the first and third groups is about 1⅔ standard deviations based on the pooled within-groups variability. The partial correlation between the contrast and its DF is:

$$\text{mult. } \hat{\eta}_{\hat{\psi}_1} = (1 - .6313)^{1/2} = .3687^{1/2} = .61$$

Based on this result, we can also say that the comparison of the first group with the third group explains about 36.9% of the variance in the DF for this contrast after removing the effect of all other sources of between-groups variability besides that due to the contrast of the first and third groups (i.e., $\hat{\psi}_1$).

The structure coefficients of *v*1 and *v*2 for the multivariate contrast $\hat{\psi}_1$ are, respectively, .26 and .84. These results indicate that (1) outcome variable *v*2 contributes relatively more to the separation of the first and third groups than *v*1 controlling for their correlation, and (2) the DF for the multivariate contrast $\hat{\psi}_1$ extracts $(.26^2 + .84^2)/2 = .387$, or 38.7% of the variance in the set of *v*1 and *v*2. Redundancy is not ordinarily computed for partial correlations, so no computations are shown for this statistic.

The author is unaware of a software program that computes approximate or exact confidence intervals for the multivariate effect size indexes described above. One hopes that this situation will change in the near future.

FOUR-GROUP DESIGN EXAMPLE

The same data set is considered in this second research example as in the first example for a two-group comparison. Briefly reviewed, Kline et al. (1993) administered the PIC to the parents of $n_1 = 199$ students in regular education and 572 students in special education. The latter group is made up of children designated as learning disabled ($n_2 = 244$), emotionally impaired ($n_3 = 132$), or cognitively impaired ($n_4 = 96$). Mean profiles of all four groups on the 12 scales of the parent-informant inventory are presented in Figure 1. As expected, cases designated as cognitively impaired or learning disabled have relatively high mean scores on scales that measure cognitive or developmental problems, and cases labeled as emotionally impaired group have relatively high mean scores on scales that indicate emotional or behavioral problems. See Appendix B for descriptive statistics for all four groups.

Insert Figure 1 about here

The results of the four-group MANOVA are:

$$\Lambda_A = .2266,\ F\ (36,\ 1938.95) = 35.17, \text{ and } p < .001$$

That the omnibus comparison for this example is statistically significant is neither surprising nor particularly informative considering the relatively large overall sample size ($N = 771$). Of greater interest are estimates of the multivariate effect size at the group and case levels. The multivariate correlation ratio equals mult. $\hat{\eta}_A^2 = 1 - .2266 = .7734$, or about 77.3% in terms of the multivariate percentage of explained variance. A variance-accounted-for effect size that may be more realistic is given by estimated tau-squared, which for this analysis with four groups and 12 outcome variables equals

$$\hat{\tau}^2 = (1 - .2266^{1/3}) = (1 - .6097) = .3903$$

which says that the overall percentage of explained variance is about 39.0%.

Table 5 reports for each of the three orthogonal discriminant functions squared canonical correlations, percentages of variance in the whole set of PIC scales explained by each function, and redundancies. Expression of estimated tau-squared as one minus the geometric mean of the complements of the squared canonical correlations is shown next (with some slight rounding error):

$$\hat{\tau}^2 = 1 - [(1 - .600)\ (1 - .332)\ (1 - .153)]^{1/3} = 1 - .2263^{1/3} = .3906$$

The three functions altogether extract a total of 59.6% of the variance of the scales. Just under half of this amount—total redundancy = .246, or 24.6%—is explained by membership in one of the four regular or special education groups through all three discriminant functions.

Insert Table 5 about here

Table 6 reports structure coefficients and $\hat{\eta}^2_A$ values for individual scales of the PIC. Because the coefficients of the three scales that measure child intellectual, developmental, or achievement status are all >.80, the first function is primarily a cognitive status dimension. The second dimension is defined mainly by scales that measure internalization or externalization. The third function does not seem all that distinct from the second except that a measure of disorganized or unusual behavior (the Psychosis scale) has a relatively high loading on it.

Insert Table 6 about here

The scatterplots presented in Figure 2 show the three functions in unstandardized form evaluated at the group means. The first function appears to separate the regular education and cognitively impaired groups, and the second function distinguishes between the emotionally impaired and the learning disabled groups. The third function appears to separate the learning disabled group from all the rest. Thus, each dimension seems to discriminate among the groups in a somewhat different way. This kind of information is lost if one neglects to evaluate the discriminant functions that underlie a one-way MANOVA.

Insert Figure 2 about here

The classification phase of a DFA was conducted to evaluate the distinctiveness of the regular and special education groups at the case level. The prior probabilities of the groups were specified as follows: regular education, .85; learning disabled, .065; emotionally impaired, .065; and cognitively impaired, .02. These are reasonably realistic group base rates. The results of this analysis based on the separate within-groups variance-covariance matrices are summarized in Table 7. The overall classification hit rate is 65.7%, which represents a 51.6% reduction in the cross-group error rate compared to chance classification. The hit rates for the regular education and cognitively impaired groups are both reasonably high (96.4 and 72.9%, respectively), but less than half of the cases in the learning disabled and emotionally impaired groups are correctly classified. Many of the cases in these two groups are incorrectly classified as either belonging to the other group or to the regular education group. It seems that parental ratings for children designated as learning disabled versus emotionally impaired are not very distinct. However, it is also difficult to differentiate these two special education categories even with results of individually-administered cognitive and achievement tests (e.g., Hallahan & Kauffman, 1977). Overall, it seems that parental ratings may play a useful role in screening for the need of some kind of special education resource (the first research example), but such ratings are not as accurate in distinguishing between specific special education categories.

Insert Table 7 about here

COMPLETELY BETWEEN-SUBJECTS FACTORIAL DESIGNS

This section assumes that all factors are fixed, not random. Contrasts in multivariate factorial designs are specified with coefficients just as in univariate designs (see Kline, 2004, chap. 7). However, the researcher typically needs a computer program for factorial MANOVA in order to compute test statistics for multivariate contrasts from which effect size indexes can be derived. For example, the following syntax for the General Linear Model (GLM) module in SPSS specifies the three multivariate simple effects of factor *A* at each level of factor *B* in a completely between-subjects 2 × 3 design with outcome variables *v*1 and *v*2:

```
GLM v1 v2 BY A B
 /LMATRIX 'A at B simple effects'
  A 1 -1 A*B 1 0 0 -1  0  0;
  A 1 -1 A*B 0 1 0  0 -1  0;
  A 1 -1 A*B 0 0 1  0  0 -1
 /DESIGN = A B A*B.
```

Each line after the LMATRIX statement specifies the comparison of A_1 with A_2 at each level of *B*. The specific syntax in other computer programs for factorial MANOVA would be different (e.g., Tabachnick & Fidell, 2001, chaps. 5-7), but the general idea is the same.

The Mahalanobis generalized distance can be computed for a multivariate contrast in a factorial design just as in a one-way design. It is derived in the two designs with the same basic formula except that the ratios of the squared contrast coefficients over the cell sizes are summed across just the cells involved in the contrast of interest (see Equation 7). (Mean difference scaling is assumed for all sets of contrast weights.) Doing so results in a generalized distance that has the same general form as in a one-way design (see Equation 6). However, the test statistic $\Lambda_{\hat{\psi}}$ in a factorial design controls for effects of all other factors in the design (e.g., Equation 8). This may not be a problem when *none* of the off-factors for a single-factor contrast vary naturally in the population to which the results should generalize (e.g., the design is experimental). The author is unaware of a method to compute standardized single-factor contrasts in multivariate factorial designs where error variance is pooled with variance due to effects of off-factors that vary naturally in the research population (see Kline, 2004, chap. 7).

The generalized correlation ratio also has the same form in a multivariate factorial design as in a multivariate one-way design. It is

$$\text{mult. } \hat{\eta}^2_{\text{effect}} = 1 - \Lambda_{\text{effect}} \qquad (14)$$

where Λ_{effect} is the Wilks's lambda test statistic for the effect of interest such as a contrast, main effect, or interaction effect. In general, mult. $\hat{\eta}^2_{\text{effect}}$ is a proportion of partial explained variance because the statistic Λ_{effect} corrects for all other effects. An exception is when mult. $\hat{\eta}^2_{\text{effect}}$ is calculated for all main and interaction effects together (e.g., mult. $\hat{\eta}^2_{A,\,B,\,AB}$ in a design with just two factors). In this case, mult. $\hat{\eta}^2_{\text{effect}}$ is interpreted as a proportion of total explained variance. An alternative variance-accounted-for effect size index for a multivariate factorial design is estimated tau-squared:

$$\hat{\tau}^2_{\text{effect}} = 1 - \Lambda^{1/r}_{\text{effect}} \qquad (15)$$

where *r* is the number of roots, the smaller of the number of degrees of freedom for the effect or the number of outcome variables. If

$r > 1$, then $\hat{\tau}^2 <$ mult. $\hat{\eta}^2$ for the same effect and thus is a more conservative descriptive measure of association. Some computer programs like GLM in SPSS actually print $\hat{\tau}^2$ instead of mult. $\hat{\eta}^2$ for effects with more than a single degree of freedom in multivariate factorial designs.

Table 8 presents a small data set for a 2 × 2 design with two outcome variables, *v*1 and *v*2. All effects are contrasts in this design (i.e., $df = 1$). The pooled within-groups correlation between the outcome variables is .418, and the two-way ANOVA error terms (MS_W) for *v*1 and *v*2 are, respectively, 5.25 and 7.50.

Insert Table 8 about here

Standardizing all univariate mean contrasts against the square roots of their respective pooled within-groups variances gives us:

	*v*1	*v*2
$g_{\hat{\psi}_A}$	.22	.91
$g_{\hat{\psi}_B}$	.22	.18
$g_{\hat{\psi}_{AB}}$	2.18	1.10

In words, the magnitudes of the univariate interaction effects are generally larger than those of the corresponding main effects in standard deviation units. Inspection of the means in Table 8 indicates that the simple effects of each factor are greater at the first level of the other factor than at the second level. The magnitude of this difference (i.e., the interaction effect) is just over two standard deviations for *v*1 and about half as large *v*2. Univariate proportions of partial variance explained by each effect are:

	*v*1	*v*2
partial $\hat{\eta}^2_A$	.02	.10
partial $\hat{\eta}^2_B$	.14	.10
partial $\hat{\eta}^2_{AB}$	.31	.24

We again see that the relative magnitudes of the univariate interaction effects are greater than those of the main effects, this time in a variance-accounted-for metric.

The dsata in Table 8 were analyzed with a computer program for factorial MANOVA. Table 9 reports the values of Wilks's lambda and effect size indexes for each multivariate main and interaction effect. There are two centroids for the multivariate main effect of factor *A* each based on scores from $n = 6$ cases. The contrast weights are $(1, -1)$, and for this effect $\Lambda_A = .8981$ (Table 9). From this information we calculate the multivariate standardized contrast as follows:

$$D_{M_{\hat{\psi}_A}} = [8\ (1 - .8981)\ (1/6 + 1/6)/.8981]^{1/2} = .55$$

The multivariate standardized contrast for the *B* main effect of .71 (Table 9) is derived in a similar way.

Insert Table 9 about here

The multivariate two-way interaction effect concerns the four cell centroids, each of which is based on $n = 3$ cases. A standard set of contrast weights for the interaction effect are

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and Λ_{AB} = .6575, so the multivariate standardized contrast for this effect is calculated as

$$D_{M_{\hat{\psi}_{AB}}} = [8\ (1 - .6575)(1/3 + 1/3 + 1/3 + 1/3)/.6575]^{1/2} = 2.36$$

which as expected is higher than that for both main effects (Table 9). The multivariate portions of partial variance explained by the various effects range from .10 for the A main effect to .34 for the two-way interaction. Because $df = 1$ for all effects in this analysis, mult. $\hat{\eta}^2 = \hat{\tau}^2$ for each effect. See Olejnik and Algina (2000) for additional examples of multivariate effect size estimation in factorial designs.

RECOMMENDED READINGS

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*, 241-286.

Tatsuoka, M. (1993). Effect size. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Vol. 1. Methodological issues* (pp. 461-479). Hillsdale, NJ: Lawrence Erbaum.

REFERENCES

Hallahan, D. P., & Kauffman, J. M. (1977). Labels, categories, behaviors: ED, LD, and EMR reconsidered. *Journal of Special Education*, *11*, 139-149.

Hand, D. J., & Taylor, C. C. (1987). *Multivariate analysis of variance and repeated measures: A practical introduction for behavioural scientists*. New York: Chapman and Hall.

Huberty, C. J. (1994). *Applied discriminant analysis*. New York: John Wiley.

Keppel, G., & Zedeck, S. (1989). *Data analysis for research designs*. New York: W. H. Freeman.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Kline, R. B., Lachar, D., & Boersma, D. C. (1993). Identification of special education needs with the Personality Inventory for Children (PIC): A hierarchical classification model. *Psychological Assessment*, *5*, 307-316.

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1988). Redundancy analysis: An alternative to canonical correlation and multivariate multiple regression in exploring interset associations. *Psychological Bulletin*, *104*, 282-289.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*, 241-286.

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, *83*, 579-586.

Schaffer, J. P., & Gillo, M. W. (1974). A multivariate extension of the correlation ratio. *Educational and Psychological Measurement*, *34*, 521-524.

Silva, A. P. D., & Stam, A. (1995). Discriminant analysis. In L. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 277-318). Washington, DC:

American Psychological Association.
Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Lawrence Erlbaum.
Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin*, *70*, 160–163.
Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.
Tatsuoka, M. (1993). Effect size. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Vol. 1. Methodological issues* (pp. 461-479). Hillsdale, NJ: Erbaum.
Thomas, D. (1992). Interpreting discriminant functions: A data analytic approach. *Multivariate Behavioral Research*, *27*, 335-362.
Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat P. D. (1984). *Multidimensional description of child personality: A manual for the Personality Inventory for Children*. Los Angeles: Western Psychological Services.

Table 1. *Raw Scores and Descriptive Statistics on Two Outcome Variables for Two Independent Samples.*

	Group			
	1		2	
	*v*1	*v*2	*v*1	*v*2
	9	12	8	10
	12	7	12	5
	13	13	11	8
	15	8	10	4
	16	10	14	8
M	13.00	10.00	11.00	7.00
s^2	7.50	6.50	5.00	6.00

Table 2. *Descriptive Statistics for Regular and Special Education Groups and Structure Correlations in the Analysis Sample.*

PIC scale		Analysis sample: Special education	Analysis sample: Regular education	r_{pb}	Structure Coefficient
	n	236	99		
ACH	M (*s*)	69.06 (10.96)	46.91 (8.14)	.705	.956
IS		78.13 (26.08)	48.38 (9.77)	.517	.701
DVL		66.73 (13.46)	45.93 (7.91)	.619	.839
SOM		53.22 (12.20)	49.34 (9.61)	.153	.208
D		60.50 (14.58)	49.00 (8.62)	.373	.506
FAM		54.27 (10.95)	46.71 (6.02)	.334	.454
DLQ		61.04 (16.67)	50.26 (8.08)	.319	.434
WDL		55.11 (12.81)	48.64 (9.43)	.241	.328
ANX		59.01 (13.09)	50.54 (9.43)	.305	.414
PSY		65.90 (19.69)	50.87 (10.06)	.367	.499
HPR		56.82 (15.32)	50.71 (10.60)	.195	.265
SSK		62.42 (13.65)	48.82 (10.44)	.438	.594

Note. PIC = Personality Inventory for Children (Wirt, Lachar, Klinedinst, & Seat, 1984). ACH = Achievement; IS = Intellectual Screening; DVL = Development; SOM = Somatic Concern; D = Depression; FAM = Family Relations; DLQ = Delinquency; WDL = Withdrawal; ANX = Anxiety; PSY = Psychosis; HPR = Hyperactivity; and SSK = Social Skills.

Table 3. *Classification Results in the Analysis and Replication Samples.*

Actual group	*n*	Predicted group: Reg	Predicted group: Spec	Hit rate (%): By group	Hit rate (%): Overall	*I* (%)
Analysis sample						
Reg	99	**95**	4	96.0	83.6	75.0
Spec	236	51	**185**	78.4		
Replication sample						
Reg	100	**90**	10	90.0	79.8	68.6
Spec	236	58	**178**	75.4		

Note. Reg = regular education, Spec = special education. Frequencies in boldface font are numbers of correct classifications.

Table 4. *Raw Scores and Descriptive Statistics on Two Outcome Variables for Three Independent Samples.*

	Group					
	1		2		3	
	*v*1	*v*2	*v*1	*v*2	*v*1	*v*2
	9	12	8	10	10	6
	12	7	12	5	11	8
	13	13	11	8	13	10
	15	8	10	4	11	5
	16	10	14	8	15	2
M	13.00	10.00	11.00	7.00	12.00	6.20
s^2	7.50	6.50	5.00	6.00	4.00	9.20

Table 5. *Proportions of Explained Variance for the Three Discriminant Functions for Scales from a Parent-Informant Measure of Child Adjustment.*

	Discriminant function			
Statistic	1	2	3	Total
Squared canonical correlation	.600	.331	.153	—
Variance extracted from the 12 PIC scales (%)	26.9	19.3	13.4	59.6
Redundancy (%)	16.1	6.4	2.1	24.6

Note. PIC = Personality Inventory for Children (Wirt, Lachar, Klinedinst, & Seat, 1984).

Table 6. *Structure Coefficients and Correlation Ratios for Scales of a Parent-Informant Measure of Child Adjustment.*

	Discriminant function			
PIC scale	1	2	3	$\hat{\eta}^2_A$
ACH	.818	.418	−.276	.471
IS	.928	−.196	.042	.530
DVL	.909	.115	−.016	.500
SOM	.115	.256	.242	.039
D	.233	.674	.494	.221
FAM	.225	.549	.280	.142
DLQ	.169	.648	.505	.195
WDL	.330	.252	.383	.109
ANX	.129	.547	.393	.133
PSY	.628	.074	.579	.290
HPR	.016	.509	.200	.092
SSK	.452	.472	.470	.230

Note. PIC = Personality Inventory for Children (Wirt, Lachar, Klinedinst, & Seat, 1984). ACH = Achievement; IS = Intellectual Screening; DVL = Development; SOM = Somatic Concern; D = Depression; FAM = Family Relations; DLQ = Delinquency; WDL = Withdrawal; ANX = Anxiety; PSY = Psychosis; HPR = Hyperactivity; and SSK = Social Skills.

Table 7. *Classification Results for the Regular and Special Education Samples.*

Actual group		Predicted group				Hit rate (%)		
	n	Reg	LD	EI	CI	By group	Overall	*I* (%)
Reg	199	**192**	3	4	0	96.4	65.7	51.6
LD	244	88	**116**	30	10	47.5		
EI	132	40	22	**63**	7	47.7		
CI	96	4	18	4	**70**	72.9		

Note. Reg = regular education; LD = learning disabled; EI = emotionally impaired; and CI = cognitively impaired. Frequencies in boldface font are numbers of correct classifications.

Table 8. *Raw Scores and Descriptive Statistics on Two Outcome Variables for a Completely Between-Subjects 2×3 Factorial Design.*

	B_1		B_2		Row means	
	*v*1	*v*2	*v*1	*v*2	*v*1	*v*2
A_1	7 8 12 9.00 (7.00)[a]	15 11 16 14.00 (7.00)	3 5 7 5.00 (4.00)	13 7 10 12.00 (9.00)	7.00	13.00
A_2	5 5 8 6.00 (3.00)	8 9 13 10.00 (7.00)	6 5 10 7.00 (7.00)	10 9 14 11.00 (7.00)	6.50	10.50
Column means	7.50	12.00	6.00	11.50	6.75	11.25

[a]Cell mean (variance)

Table 9. *Multivariate Effect Size Indexes for the Data in Table 8.*

Source	Λ	$D_{M_{\hat{\psi}}}$	mult. $\hat{\eta}^2$
A	.8981	.55	.10
B	.8421	.71	.16
AB	.6575	2.36	.34

Note. Results for mult. $\hat{\eta}^2$ are proportions of partial variance.

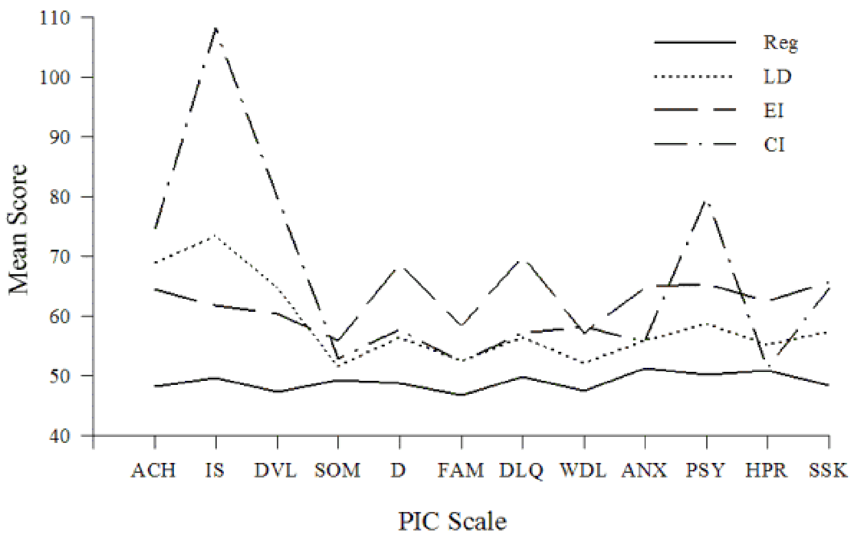

Figure 1. Mean profiles on the Personality Inventory for Children (PIC; Wirt, Lachar, Klinedinst, & Seat, 1984) for students in regular education (Reg) and those designated as learning disabled (LD), emotionally impaired (EI), or cognitively impaired (CI). ACH = Achievement; IS = Intellectual Screening; DVL = Development; SOM = Somatic Concern; D = Depression; FAM = Family Relations; DLQ = Delinquency; WDL = Withdrawal; ANX = Anxiety; PSY = Psychosis; HPR = Hyperactivity; and SSK = Social Skills.

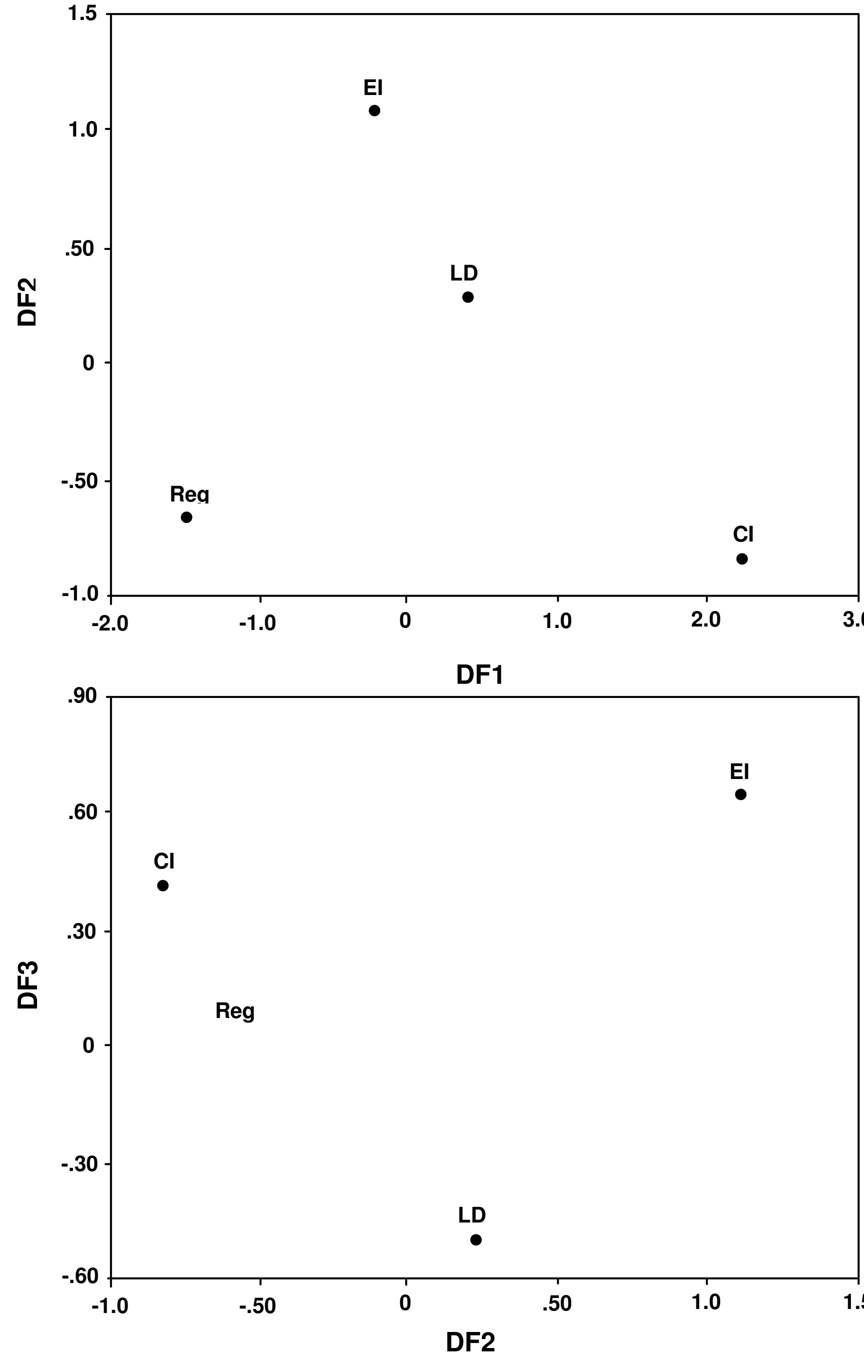

Figure 2. Unstandardized discriminant functions evaluated at group means. Reg = regular education; LD = learning disabled; EI = emotionally impaired; and CI = cognitively impaired.

APPENDIX A: DESCRIPTIVE STATISTICS, EXAMPLE 1

Analysis sample, regular education, n = 99

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	46.91	8.14	1.000											
IS	48.38	9.77	.131	1.000										
DVL	45.93	7.91	.851	.218	1.000									
SOM	49.34	9.61	.219	.089	.096	1.000								
D	49.00	8.62	.157	−.043	.138	.307	1.000							
FAM	46.71	6.02	.259	−.166	.149	.324	.467	1.000						
DLQ	50.26	8.08	.216	−.238	.114	.088	.439	.370	1.000					
WDL	48.64	9.43	.043	−.035	−.013	.327	.615	.302	.166	1.000				
ANX	50.54	9.43	.093	.009	.081	.203	.809	.326	.287	.407	1.000			
PSY	50.87	10.06	.298	−.001	.253	.369	.645	.314	.405	.559	.360	1.000		
HPR	50.71	10.60	.045	−.026	.113	.062	−.090	−.141	.284	−.425	−.103	.045	1.000	
SSK	48.82	10.44	.448	−.210	.444	.290	.524	.324	.413	.256	.362	.449	.189	1.000

Analysis sample, special education, n = 236

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	69.06	10.96	1.000											
IS	78.13	26.08	.495	1.000										
DVL	66.73	13.46	.795	.703	1.000									
SOM	53.22	12.20	.141	.088	.098	1.000								
D	6(	14.58	.266	−.093	.117	.427	1.000							
FAM	54.27	10.95	.230	−.045	.100	.242	.483	1.000						
DLQ	61.04	16.67	.216	−.137	.054	.388	.653	.432	1.000					
WDL	55.11	12.81	.287	.122	.236	.281	.644	.301	.366	1.000				
ANX	59.01	13.09	.216	−.058	.070	.467	.826	.417	.525	.422	1.000			
PSY	65.90	19.69	.371	.403	.495	.328	.549	.278	.374	.551	.389	1.000		
HPR	56.82	15.32	-.002	−.157	−.086	.223	.236	.138	.457	−.076	.217	.234	1.000	
SSK	62.42	13.65	.297	.131	.291	.159	.561	.330	.427	.452	.377	.710	.414	1.000

Replication sample, regular education, n = 100

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	49.50	10.24	1.000											
IS	50.81	9.56	.288	1.000										
DVL	48.76	8.80	.849	.315	1.000									
SOM	49.15	7.36	.151	.214	.120	1.000								
D	48.55	8.52	.178	.105	.138	.291	1.000							
FAM	46.92	6.12	.053	−.132	.049	.108	.269	1.000						
DLQ	49.50	8.34	.254	.132	.078	.211	.390	.497	1.000					
WDL	46.46	8.06	.255	.246	.204	.169	.564	.006	.179	1.000				
ANX	52.01	8.80	.120	.041	.049	.325	.755	.104	.273	.262	1.000			
PSY	49.49	8.68	.261	.243	.252	.325	.528	.141	.138	.529	.269	1.000		
HPR	51.15	9.04	.078	−.106	.004	−.026	−.295	.186	.131	−.397	−.189	−.177	1.000	
SSK	48.06	9.18	.465	.038	.392	.148	.593	.265	.349	.415	.352	.583	−.027	1.000

Replication sample, special education, n = 236

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	68.59	11.90	1.000											
IS	76.30	25.82	.452	1.000										
DVL	66.61	13.11	.739	.722	1.000									
SOM	53.04	13.80	.070	−.062	−.014	1.000								
D	59.78	16.14	.119	−.077	.010	.460	1.000							
FAM	54.07	11.81	.045	−.152	.001	.276	.406	1.000						
DLQ	59.85	17.58	.181	−.108	.054	.504	.621	.508	1.000					
WDL	54.42	12.84	.245	.191	.245	.266	.592	.144	.327	1.000				
ANX	57.81	14.34	.032	−.049	−.029	.414	.874	.382	.532	.416	1.000			
PSY	63.92	17.52	.273	.435	.451	.278	.541	.201	.376	.606	.439	1.000		
HPR	56.00	14.96	.038	−.205	−.109	.170	.199	.187	.535	−.204	.180	.073	1.000	
SSK	59.98	13.76	.219	.051	.227	.274	.649	.257	.477	.499	.502	.674	.335	1.000

Note. Scales of the Personality Inventory for Children (Wirt, Lachar, Klinedinst, & Seat, 1984): ACH = Achievement; IS = Intellectual Screening; DVL = Development; SOM = Somatic Concern; D = Depression; FAM = Family Relations; DLQ = Delinquency; WDL = Withdrawal; ANX = Anxiety; PSY = Psychosis; HPR = Hyperactivity; and SSK = Social Skills.

APPENDIX B: DESCRIPTIVE STATISTICS, EXAMPLE 2

Regular education sample, n = 199

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	48.21	9.32	1.000											
IS	49.60	9.72	.230	1.000										
DVL	47.35	8.47	.852	.283	1.000									
SOM	49.25	8.53	.177	.139	.102	1.000								
D	48.77	8.55	.162	.026	.131	.298	1.000							
FAM	46.81	6.06	.143	−.145	.097	.226	.367	1.000						
DLQ	49.88	8.20	.228	−.057	.085	.141	.415	.434	1.000					
WDL	47.54	8.81	.131	.076	.070	.263	.588	.159	.175	1.000				
ANX	51.28	9.12	.116	.035	.077	.251	.778	.218	.275	.328	1.000			
PSY	50.18	9.39	.260	.101	.234	.350	.589	.230	.280	.549	.311	1.000		
HPR	50.93	9.83	.064	−.059	.062	.027	−.185	.012	.210	−.413	−.139	−.051	1.000	
SSK	48.44	9.81	.440	−.099	.402	.232	.555	.294	.382	.327	.353	.508	.095	1.000

Learning disabled sample, n = 244

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	68.94	10.83	1.000											
IS	73.34	18.74	.380	1.000										
DVL	64.86	10.57	.763	.567	1.000									
SOM	51.66	11.34	.168	.137	.137	1.000								
D	56.46	12.81	.323	.192	.245	.417	1.000							
FAM	52.52	10.55	.235	.075	.195	.252	.390	1.000						
DLQ	56.55	14.31	.313	.154	.280	.438	.565	.418	1.000					
WDL	52.12	12.32	.278	.234	.275	.276	.661	.203	.389	1.000				
ANX	55.93	12.15	.201	.166	.160	.382	.830	.354	.470	.415	1.000			
PSY	58.76	14.42	.289	.291	.392	.340	.691	.318	.548	.607	.501	1.000		
HPR	55.16	14.08	.034	−.110	−.021	.147	.094	.165	.400	−.229	.093	.169	1.000	
SSK	57.41	12.78	.289	.068	.263	.135	.584	.289	.440	.406	.401	.711	.333	1.000

Emotionally impaired sample, n = 132

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	64.46	12.76	1.000											
IS	61.80	24.39	.413	1.000										
DVL	60.48	11.71	.805	.596	1.000									
SOM	55.98	15.08	.166	.032	.108	1.000								
D	68.73	18.20	.308	−.116	.148	.417	1.000							
FAM	58.36	13.05	.188	−.062	.143	.234	.447	1.000						
DLQ	70.06	20.83	.345	−.103	.196	.458	.618	.455	1.000					
WDL	57.17	13.50	.360	.142	.313	.293	.597	.220	.304	1.000				
ANX	64.93	15.77	.244	−.003	.117	.399	.855	.372	.485	.468	1.000			
PSY	65.35	18.74	.276	.288	.336	.386	.593	.220	.337	.607	.493	1.000		
HPR	62.50	16.23	.161	.019	.135	.213	.173	.015	.455	−.096	.151	.277	1.000	
SSK	65.73	14.33	.304	.099	.299	.323	.594	.230	.403	.496	.450	.725	.419	1.000

Cognitively impaired sample, n = 96

	M	*SD*	ACH	IS	DVL	SOM	D	FAM	DLQ	WDL	ANX	PSY	HPR	SSK
ACH	74.53	8.00	1.000											
IS	108.25	16.59	.364	1.000										
DVL	79.79	12.74	.719	.618	1.000									
SOM	52.97	13.40	−.031	−.109	−.087	1.000								
D	57.66	12.26	.169	.061	.151	.487	1.000							
FAM	52.59	9.40	.095	−.215	.001	.218	.272	1.000						
DLQ	57.11	12.35	.203	−.134	−.028	.389	.535	.403	1.000					
WDL	58.19	11.70	.086	−.129	.037	.163	.606	.166	.252	1.000				
ANX	55.74	11.49	.119	.065	.102	.562	.814	.314	.473	.299	1.000			
PSY	79.95	19.53	.330	.300	.410	.176	.580	.268	.439	.436	.428	1.000		
HPR	51.23	13.49	.227	−.023	−.050	.158	.232	.182	.649	−.082	.214	.287	1.000	
SSK	64.63	12.58	.296	.135	.308	.098	.618	.253	.454	.457	.415	.657	.396	1.000

Note. Scales of the Personality Inventory for Children (Wirt, Lachar, Klinedinst, & Seat, 1984): ACH = Achievement; IS = Intellectual Screening; DVL = Development; SOM = Somatic Concern; D = Depression; FAM = Family Relations; DLQ = Delinquency; WDL = Withdrawal; ANX = Anxiety; PSY = Psychosis; HPR = Hyperactivity; and SSK = Social Skills.