Gifted Child Quarterly
2020, Vol. 64(1) 41–55

DOI: 10.1177/0016986219890352
journals.sagepub.com/home/gcq
SAGE

Making Meaning out of MANOVA: The Need for Multivariate Post Hoc Testing in Gifted Education Research

Kendal N. Smith[1], **Kristen N. Lamb**[2], and **Robin K. Henson**[1]

Abstract

Multivariate analysis of variance (MANOVA) is a statistical method used to examine group differences on multiple outcomes. This article reports results of a review of MANOVA in gifted education journals between 2011 and 2017 ($N = 56$). Findings suggest a number of conceptual and procedural misunderstandings about the nature of MANOVA and its application, including pervasive use of univariate post hoc tests to interpret MANOVA results. Accordingly, this article aims to make MANOVA more accessible to gifted education scholars by clarifying its purpose and introducing descriptive discriminant analysis as a more appropriate post hoc technique. A heuristic data set is used to demonstrate the procedures for running a descriptive discriminant analysis, both in place of a one-way MANOVA and as a post hoc analysis to a factorial design. SPSS and R syntax are provided.

Keywords

MANOVA, discriminant analysis, multivariate analysis

In a comprehensive review of statistical methods used in gifted education journals between 2006 and 2010, Warne, Lazo, Ramos, and Ritter (2012) found multivariate analysis of variance (MANOVA) to be one of the most commonly employed procedures, reported in 23% of quantitative articles. An extension of univariate analysis of variance (ANOVA), MANOVA allows researchers to identify group differences on a combination of outcome variables simultaneously, rather than on a single measure (Huberty & Morris, 1989). If differences are found, MANOVA results (like ANOVA) are typically followed by a post hoc procedure to aid in substantive interpretation (Field, 2013). MANOVA is unique from ANOVA, though, in that these post hoc procedures must address not only (a) what groups are different but also (b) what outcome variables matter in creating those differences.

In their review, however, Warne et al. (2012) observed most gifted education scholars used inappropriate post hoc methods, creating "serious doubts about whether many researchers understand how to interpret such statistical tests" (p. 145). Indeed, MANOVA results were most frequently followed with a series of univariate tests, a method that has been variously described as "illogical" (Zientek & Thompson, 2009, p. 345), "conceptually inconsistent" (Barton, Yeatts, Henson, & Martin, 2016, p. 366), "irrelevant" (Huberty & Morris, 1989, p. 307), a "bad habit" (Grice & Iwasaki, 2007, p. 200), and something "researchers should completely avoid" (Warne, 2014, p. 5).

Following a MANOVA with a series of ANOVAs suggests a fundamental misunderstanding about the natures of these analyses. Although both investigate mean differences between two or more groups, ANOVA effects are calculated from differences in observed scores on a single outcome variable, whereas MANOVA effects emerge from composite (or synthetic) scores on a collection of outcome variables (Huberty & Olejnik, 2006). Importantly, MANOVA does *not* analyze outcome variables individually—a process equivalent to conducting multiple ANOVAs—but simultaneously considers the relationships between the outcomes by linearly combining them to generate one or more synthetic or composite variables (Sherry, 2006). These composites represent unobserved constructs that can best account for the group differences, and thus become the focus of the analysis (Kieffer, Reese, & Thompson, 2001).

As such, ANOVA and MANOVA do not analyze the same variables and, thereby, address different research questions (Enders, 2003; Zientek & Thompson, 2009). To illustrate this point, several methodologists have demonstrated how analysis of group differences in the same data set can produce

[1]University of North Texas, Denton, TX, USA
[2]University of Washington, Seattle, WA, USA

Corresponding Author:
Kendal N. Smith, University of North Texas, 1300 W. Highland Street, Denton, TX 76201, USA.
Email: kendal.smith@unt.edu

vastly discrepant results, depending on whether a series of ANOVAs or a single MANOVA was conducted (Fish, 1988; Huberty & Morris, 1989; Maxwell, 1992; Thompson, 1999; Zientek & Thompson, 2009).

Because constructs of interest to social scientists rarely exist in a vacuum, multivariate analyses often afford a more nuanced picture that better reflects the complexity of variable relationships typical of the social world (Fish, 1988; Zientek & Thompson, 2009). As noted by Sherry and Henson (2005), "Determining outcomes based on research that separately examines singular causes and effects may distort the complex reality of human behavior and cognition" (p. 38). Yet the advantage of capturing this complexity is lost if a multivariate effect is interpreted with univariate post hocs which, by definition, ignore outcome variable relationships. It makes little sense to first declare an interest in an omnibus multivariate result and then try to interpret that result univariately. If a multivariate analysis is used, it should be accompanied by a multivariate interpretation.

The most frequently recommended and simplest multivariate post hoc procedure to MANOVA is descriptive discriminant analysis (DDA; Enders, 2003; Field, 2013; Huberty & Olejnik, 2006; Warne, 2014). DDA enables explanation of group differences in a multivariate context by providing information about which groups differ on the composite as well as which outcome variables are responsible for its creation. Unfortunately, DDA is rarely used and widely misunderstood.

Huberty and Morris (1989), for example, found only 2 of 222 articles published in prominent psychology journals used DDA to interpret group differences on multiple outcomes, whereas 98% used univariate procedures. Subsequent reviews by Keselman et al. (1998) and Kieffer, Reese, and Thompson (2001) revealed little change in the prevalence of univariate post hoc testing, even though the "methodological literature appears to be unequivocally opposed to this practice" (Enders, 2003, p. 41). More recent reviews indicate univariate tests remain pervasive (Barton et al., 2016; Tonidandel & LeBreton, 2013; Warne, 2014), and the gifted education literature is no exception (Warne et al., 2012).

Purpose of the Present Article

Given the potential contributions of multivariate analysis in the gifted education literature and in light of the continued use of univariate post hocs with MANOVA, the present article has four primary aims. First, we review the recent gifted education research to document whether the issue remains widespread. Second, this review also allows comment on whether some broader statistical approaches used by gifted education researchers reflect best methodological practices. Third, we attempt to bring conceptual clarity to MANOVA by framing it within the context of related statistical analyses familiar to researchers and identifying the types of research scenarios in which it can (and cannot) be meaningfully applied. Finally, we provide a heuristic demonstration of how to employ DDA as a multivariate post hoc to MANOVA for both main and interaction effects. Accordingly, this article aims to make MANOVA more accessible to gifted education scholars by clarifying its purpose and illustrating how DDA can be used to interpret MANOVA effects. We use a small-sample, heuristic data set to demonstrate the procedures for conducting a DDA, both in place of a one-way MANOVA and as a post hoc analysis to a factorial design.

Table 1. MANOVA Post Hoc Procedures Reported in Gifted Education Journal Articles, 2011-2017 (N = 51).

	Post hoc method			
	Only multivariate	Only univariate	Both multivariate and univariate	Total
Year				
2011	1	13	3	17 (33.3%)
2012	0	6	2	8 (15.7%)
2013	0	2	0	2 (3.9%)
2014	0	5	0	5 (9.8%)
2015	0	4	0	4 (7.8%)
2016	0	5	1	6 (11.8%)
2017	1	7	1	9 (17.7%)
Journal				
GCQ	1	7	2	10 (19.6%)
HAS	1	9	0	10 (19.6%)
JAA	0	7	1	8 (15.7%)
JEG	0	8	3	11 (21.6%)
RR	0	11	1	12 (23.5%)
Total	2 (3.9 %)	42 (82.4%)	7 (13.7%)	

Note. GCQ = Gifted Child Quarterly; HAS = High Ability Studies; JAA = Journal of Advanced Academics; JEG = Journal for the Education of the Gifted; RR = Roeper Review. All multivariate post hocs were some form of discriminant analysis.

MANOVA in Gifted Education Research

We reviewed the same five gifted education journals as Warne et al. (2012) to see whether MANOVA post hoc practices have improved. Results are presented in Table 1 by journal and year. Out of 303 articles published from 2011 to 2017 that reported one or more quantitative analyses, 56 (18.5%) included results of one or more MANOVAs. Four of these reported results of intervention studies, one that used a randomized controlled design, two that used pretest/posttest designs, and one that used a posttest only design. Three additional articles used repeated measures MANOVA to examine within-subject differences over time. The remaining 49 articles reported observational analyses using one or more between-subject factors to investigate differences on multiple outcomes. Unsurprisingly, the most common grouping variable was some form of gifted status or participation, followed by gender, ethnicity, and teacher and parent comparisons.

Five of the 56 articles reported nonstatistically significant and/or negligible effects and authors did not conduct post hoc analysis. Of the remaining 51 articles, 49 (96.1%) reported univariate post hoc testing to at least one MANOVA, indicating this approach continues to predominate. In three articles, authors conducted both univariate and multivariate post hoc testing of the same MANOVA, and in four articles, authors reported univariate post hoc results for one or more MANOVAs but multivariate for one or more others. Authors of only two (3.9%) articles reported solely multivariate post hoc analysis and interpretation (Chae & Gentry, 2011; Puryear & Kettler, 2017).

Our review also revealed a number of practices reflecting misunderstandings of the purpose, capabilities, and interpretation of MANOVA in the gifted education literature. These misunderstandings provide context that may help illustrate why multivariate post hocs are not more frequently employed.

Rationale for Using MANOVA

In 43 (76.8%) of the 56 total articles, authors failed to explain why MANOVA was deemed appropriate for answering the stated research question(s), a basic reporting requirement for any study (American Psychological Association [APA], 2010; see also Appelbaum et al., 2018). For those that did, only two (3.6%) articles explicitly linked the multivariate analysis to the multivariate nature of the studied outcomes.

Authors of six (10.7%) articles stated MANOVA was used to protect against Type 1 error inflation. However, though it is true that simultaneous analysis of multiple outcome variables can limit the chance of making a Type 1 error anywhere in the study, this benefit is nullified if the MANOVA is followed by successive ANOVAs. Rather, "the comfort of statistical protection is an illusion" (Huberty & Morris, 1989, p. 307), and univariate tests render the incorrectly assumed Type 1 error protection of MANOVA irrelevant. The family-wise error rate still increases, potentially leading researchers to find nonexistent effects. Bonferroni adjustments, applied in 18 (32.1%) of the reviewed articles, can also be problematic because they tend to overcorrect the problem, decreasing power and inflating the Type 2 error rate (Warne, 2014). The What Works Clearinghouse instead recommends the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) for controlling the false discovery rate, though these statistical maneuvers again ignore the multivariate nature of the derived effect.

Of course, if the variables are not related, multivariate analyses simplify to their univariate counterparts and are superfluous. Although true orthogonal outcomes are unlikely in the social sciences (Meehl, 1990), it might still be preferable to model them that way if the outcome relationships are not relevant to the research purpose. For example, researchers conducting intervention studies with potential policy implications might be interested in gains on any one of several achievement outcomes without regard to the others. In these cases, appropriate error control techniques should be applied directly to the tests of the univariate outcomes. As mentioned, however, in the majority of the reviewed articles authors used MANOVA to describe group differences on a set of conceptually related individual difference variables such as types of self-concept, creativity, perfectionism, thinking styles, emotional intelligence, approaches to learning, motivational constructs, and parent or teacher characteristics. In studies like these, a multivariate approach can provide a more comprehensive picture of the variable relationships and how they differ between groups.

Assumptions, Sample Size, and Correlation Matrices

Authors of only five (8.9%) articles discussed whether the statistical assumptions for MANOVA were satisfied in their data, such as multivariate normality and homogeneity of within-group variance/covariance matrices. In 36 (64.3%) articles, authors failed to address these issues even at the univariate level. MANOVA hypothesis tests can be questionable when these assumptions are not met, particularly with small samples or unequal group sizes (Finch & French, 2013). Total sample size in the reviewed articles ranged from $N = 46$ to $N = 5{,}200$ (median = 216). Authors of 19 (33.9%) articles did not specify group sample sizes or had discrepancies between group sizes reported in the text and tables, but of those that were reported, group sizes ranged from $n = 7$ to $n = 3{,}097$. The largest and smallest groups differed by a factor less than 1.25 in 16 articles, between 1.25 and 2.00 in 9 articles, and greater than 2.00 in 12 articles. Excluding two articles in which authors analyzed secondary data sets, no authors included a sample size justification or a priori power analysis.

Half of the reviewed articles also failed to include correlations of the outcome variables. Because the underlying relationships between outcome measures should be of primary interest when using any multivariate technique, these correlations should always be reported. If authors are unable to make the data available, published correlation matrices also enable other researchers to run secondary analyses that may yield fruitful supplementary information (Nimon, 2015; Zientek & Thompson, 2009). However, these correlations should correspond to the sample actually included in the analysis; a handful of authors reported correlations for a study's full sample but included only a subset of that sample in the MANOVA. Most reported correlations ranged from .30 to .70, though a few articles included outcomes correlated above .90. Relationships this strong suggest redundancy among the variables and generally leads to a loss of power if included in the same analysis (Finch & French, 2013).

Design Issues

We also identified substantial confusion over the basic design of a MANOVA and its capabilities. For example, in six articles authors stated that they conducted a MANOVA, but additional references to the inclusion of covariates suggested the analysis conducted was actually a multivariate analysis of *co*variance; in one article, authors reported interaction effects in what was cited as a one-way design; in another, authors reported a factorial design but results indicated a series of one-way MANOVAs were run; and in two articles, authors appeared to mistake the number of dependent variables for the number of levels in a grouping variable. Additionally, it was often unclear which variables were included in a given analysis, such as in articles that reported results of multiple MANOVAs using different grouping variables and/or different subsets of outcome variables. In total, the design of one or more reported MANOVAs was either ambiguous or inaccurately stated in 23 (41.1%) articles.

Statistical Significance and Effect Size Reporting and Interpretation

The lack of clarity in MANOVA designs extended to reported MANOVA results. For example, authors of 19 (33.9%) articles did not report the multivariate test statistic used to estimate the reported F statistic in multivariate hypothesis tests, and an additional 6 (10.7%) identified the multivariate statistic but did not include its actual value. Because the MANOVA output of most statistical software typically includes results of four different statistical significance tests (discussed below) that do not always produce the same results, this information is important to include. Twenty-seven (48.2%) articles also did not include exact p values as expected by APA guidelines, an issue similarly noted in Warne et al.'s (2012) broader review. In interpreting these tests, authors of two articles took the nonstatistically significant result of a multivariate test as evidence of no group differences on each observed outcome in the analysis, suggesting not only misunderstandings about null hypothesis testing more generally but again indicating a lack of awareness about the difference between observed outcome variables and the multivariate composite generated in a MANOVA.

Surprisingly, authors of 14 (25%) articles failed to report any type of effect size. An additional 10 (17.9%) articles only included univariate effect sizes or did not provide enough information to determine if an effect size was multivariate. Of the 32 articles that did specify multivariate effects, 87.5% reported partial eta squared (η^2_p), likely a consequence of its inclusion in statistical output. However, in the majority of these articles, authors either made no attempt to interpret the multivariate effect size statistics (75%) or limited interpretation to "small," "medium," and "large" (12.5%) without explanation or mention of prior literature, theoretical rationale, or practical meaning. Typically, the multivariate effect size was ignored altogether, and researchers relied solely on statistical significance as justification for subsequent univariate tests.

Nine of the 56 (16.1%) total articles proceeded to post hoc analysis despite finding negligible effects and nonstatistically significant results in the MANOVA. As Barton et al. (2016) noted, this amounts to interpreting "the nothingness that was found" (p. 372). Conversely, no authors attempted to interpret a multivariate interaction effect, despite 16 (28.6%) articles that reported results suggesting interpretation may have been warranted. These findings are unfortunate because effect size reporting and interpretation should be expected as standard practice today (APA, 2010; Appelbaum et al., 2018).

Multivariate Interpretation Errors

In articles that did report a multivariate post hoc, it appeared the method remained misunderstood. For instance, three articles reported both univariate and multivariate post hoc analyses. The authors relied on the univariate tests to determine which groups differed and the multivariate to assess how each dependent variable contributed to the effect. As previously discussed, however, these procedures analyze different variables (Zientek & Thompson, 2009). Consequently, these articles described group differences on the observed variables and then interpreted the source of those differences in terms of the composite variables.

We also noted inaccuracies in the multivariate interpretations, such as authors of two articles equating the standardized function coefficients and structure coefficients generated by a DDA analysis. These are different things, the former referring to the standardized weights used to create the composite variable and the latter signifying the correlations between the observed outcomes and composite variable (Henson, 2002). Two articles also reported structure coefficients greater than one, an impossible result given that structure coefficients are simply correlations.

Overall, our findings support Warne et al.'s (2012) suspicion that many gifted education researchers do not understand the purpose of MANOVA or how to interpret MANOVA results. Grice and Iwasaki (2007) attributed the persistent use of univariate post hoc tests to "a paucity of clear examples that demonstrate appropriate procedures" (p. 200). Our review confirms continued need for such demonstration in the gifted education literature. Alternatively, Huang (2020) argued that the prevalence of univariate post hoc analysis implies most researchers using MANOVA are actually interested in univariate questions. Although we found little support for this inference in our review, the general lack of discussion connecting research question(s) to method(s) makes it difficult to draw firm conclusions. Regardless, our purpose is to provide a primer for a more appropriate use when there is indeed congruence between research question and method.

DDA as a MANOVA Post Hoc

MANOVA and its appropriate post hoc procedures are best viewed through the lens of the broader general linear model (GLM), the "single analytic family" (Thompson, 2006, p. 360) that guides all classical statistical analyses. GLM analyses are correlational in nature and generate r^2-type effect sizes by applying weights to observed scores to create synthetic or composite variables that maximize explained variance. The GLM is also hierarchical, such that simpler analyses are really just subsets or special cases of others (Cohen, 1968; Graham, 2008; Knapp, 1978). Thus, one-way ANOVA can be thought of as a correlation between a grouping variable and an outcome variable, making it a special type of multiple regression. Likewise, one-way MANOVA can be thought of as a correlation between a grouping variable and a set of outcome variables, making it a special type of canonical correlation. All these analyses, in turn, are special types of path or structural equation models (Graham, 2008).

Although its usefulness is often obscured by inconsistent terminology and the idiosyncrasies of statistical software packages, the familial nature of the GLM has important interpretative implications. For instance, Thompson (2015) noted that, for any GLM analysis, the researcher is ultimately concerned with two questions: (a) Is there anything worth interpreting? (b) If so, where is it coming from? To answer the first question, researchers may look to a combination of effect sizes, interval estimates, statistical significance tests, or other results, all of which should be informed by the context of the current study and prior literature. If no effects are judged to be of practical significance and worth interpreting, then the second question is irrelevant and the interpretation is over (though this finding itself may be quite meaningful). For any noteworthy effect of interest, however, deeper investigation of underlying variable relationships is warranted to determine the nature and origin of the effect.

In MANOVA, an affirmative answer to question one means group membership can account for a meaningful portion of the variance in some linear combination of the outcome variables (Huberty & Morris, 1989). The second question is a matter of determining the underlying structure of that linear combination (composite variable) and if there are more than two groups, determining which groups vary on the composite variable and to what extent (Field, 2013).

Because these composite variables are frequently "more than the sum of their parts" (Thompson, 1999, p. 20), they cannot be reliably deconstructed with multiple univariate ANOVAs. Attempts to do so create mismatches between univariate answers to the multivariate research questions. Instead, the composite is analogous to the synthetic predicted dependent variable scores ($\hat{Y}$) in multiple regression; only in MANOVA, it is the outcome variables rather than the predictor variables that are linearly combined to create the effect. To interpret from where the R^2 effect in regression is coming, researchers can consult the standardized weights (β) used in the linear equation as well as the correlations between each predictor variable and the $\hat{Y}$ composite, called structure coefficients (r_s; Courville & Thompson, 2001). These same pieces of information are needed to interpret a MANOVA effect, except the pieces are applicable for the outcome variables (Enders, 2003).

Fortunately, this is exactly the information DDA provides. DDA is the most appropriate technique for evaluating MANOVA effects, namely because it is mathematically equivalent to a one-way MANOVA (Huberty & Olejnik, 2006). Like MANOVA, DDA creates a composite variable (from the observed dependent variables) that maximizes between-group differences relative to within-group differences, and thus yields an overall variance-accounted-for effect size (e.g., multivariate η^2). The particular advantage of DDA, however, is that it also generates the standardized weights and structure coefficients for each dependent variable, which enables meaningful interpretation of the composite variable in multivariate context. Hence, if there is only one grouping variable in the analysis, DDA can answer the two interpretive questions noted above and replace the use of MANOVA altogether, with no need for post hoc tests (Sherry, 2006). For two or more grouping variables, a factorial MANOVA can be run with DDA as a subcommand to provide the same information to interpret either main effects or an interaction effect (Enders, 2003).

Descriptive Versus Predictive Discriminant Analysis

Before demonstrating how to run and interpret a DDA, it is helpful to distinguish it from a related technique called predictive discriminant analysis (PDA). Like many statistical concepts, DDA suffers from inconsistent nomenclature in the literature, increasing confusion over its purpose and application. Much of this ambiguity stems from failure to differentiate between DDA and PDA. PDA assesses the ability of a set of response variables to classify a sample into groups, such as how well a set of screening measures can predict gifted identification (e.g., Pyryt, 2004). DDA and PDA are typically lumped together under a number of terms, including *discriminant function analysis, linear discriminant analysis, multiple discriminant analysis, classification analysis, canonical discriminate analysis*, or simply *discriminant analysis* (Huberty & Hussein, 2003). However, these analyses are distinct, with some of these terms more appropriately applied to one or the other. Both combine a set of variables to form composite functions, but in PDA, the grouping variable is the criterion, and the purpose is to derive a rule from the discriminant function scores to optimally predict or classify group membership (Boedeker & Kearns, 2019; Huberty & Olejnik, 2006), similar to logistic regression. Unlike DDA, PDA is not part of the GLM because adding predictor variables can actually decrease the classification accuracy (Thompson, 2015). Practically speaking, this means that the most important information for PDA interpretation (the classification hit rate) is irrelevant for interpreting notable MANOVA or DDA effects, which emphasize the structure of the composite variables that best separate the groups.

Statistical software packages sometimes include DDA and PDA results in the same output, leading applied researchers to report both, regardless of the purpose of the analysis. We noted this tendency in our review of MANOVA in the gifted education literature. Two of the seven articles that conducted some form of multivariate post hoc reported a mixture of DDA and PDA results without explanation. We recommend researchers use the descriptive/predictive label to clearly delineate the focus of their analysis and only report and interpret the applicable results, as demonstrated below for DDA.

Heuristic Demonstration

In this section, we demonstrate the procedures for running a DDA, first in place of a one-way MANOVA and then as a post hoc to a factorial MANOVA. We describe some of the SPSS procedures in more detail below, but the data and code to reproduce the analyses in SPSS (v.24) and in R (v.3.6.1, R Core Team, 2019) are available on the Open Science Framework (https://osf.io/vyxgt/).

We have chosen to use a small-sample, heuristic data set to walk through the most important considerations in these analyses but refer readers to Barton et al. (2016), Enders (2003), Sherry (2006), or Warne (2014) for tutorials using real data sets, and to Puryear and Kettler (2017) for a real-world example from the gifted education literature. In line with the majority of studies in our review, our examples describe between-subject differences on a set of related outcome variables. Although our goal is to illustrate a method for interpreting MANOVA effects in a manner consistent with their multivariate nature, we remind readers that, in practice, the decision to use MANOVA cannot be separated from broader considerations of design, measurement, and context.

DDA in Lieu of One-Way MANOVA

The data set includes the hypothetical scores of 80 college students on three outcome variables: mathematics self-concept, fear of failure, and social competence. The question under investigation is whether differences exist between humanities (n = 40) and STEM (n = 40) majors on these three outcomes. We have chosen a multivariate analysis because we are interested in the possible differences between the majors on all three outcomes simultaneously. Because there is only one grouping variable, DDA will generate all the information required to identify and interpret notable effects, with no need for a prior MANOVA. We employ the two-question strategy outlined previously to frame our interpretation.

SPSS Procedures. DDA can be found in the SPSS drop-down menu under Analyze > Classify > Discriminant. (To reiterate the point from above, although the "Classify" subheading is really applicable to PDA and not DDA, SPSS includes results for both analyses in the same output.) From there, move the grouping and outcome variables into the appropriate boxes, and request all descriptive statistics and plots from the boxes on the right side. It is also important to click on the button labeled "Save" and then on "Discriminant Scores." This will save the discriminant function scores resulting from the equation in the analysis as a new composite variable in the data set.

Descriptives and Assumptions. The first part of the DDA output displays the means and standard deviations for each group on the outcome variables. Along with the variable correlations, these descriptives should always be reported (APA, 2010) and for the current example are listed in Table 2.

As with all statistical procedures, certain assumptions must be met to interpret the analysis with confidence. Because DDA is equivalent to a one-way MANOVA, the same assumptions apply to both. Most important, the residuals should be independent and follow a multivariate normal distribution, and the within-group variance/covariance matrices should be approximately equal (Sherry, 2006). Because there are no definitive tests or rules about when assumptions do and do not hold, we advise using (and reporting) multiple sources of evidence to make these determinations, especially if the data set cannot be made available. Moreover, we encourage researchers to specify as many of these decisions as possible before examining their data to limit the influence of analytic flexibility (Leys, Delacre, Mora, Läkens, & Ley, 2019; Simmons, Nelson, & Simonsohn, 2011).

The independence assumption implies that the predictor variables account for all the systematic variance in the outcomes; everything left over is random error. With college major as the only predictor in our example, this means that conditional on that variable, the student scores should be unrelated. Although plausible if the data came from a truly random sample, in education research this is rarely the case, and data collected in school contexts often exhibit some degree of dependence because observations of the same students or of students in the same classroom or school tend to be related to each other (McCoach & Adelson, 2010). Ignoring these relationships can bias estimates of the standard errors and sometimes the coefficients (Huang, 2018). There are a number of ways to address clustering, depending primarily on the number of clusters in the data and whether the cluster effects are of explicit interest or the researcher wants only to correct for their influence. We do not address these issues here but refer readers to McCoach (2010) for an introduction to the multilevel modeling approach and to McNeish, Stapleton, and Silverman (2017) for a comparison of several alternatives.

Multivariate normality can be graphically assessed by examining plots of the ordered Mahalanobis distances (D^2) and paired chi-squared values (χ^2) of the model residuals. If the data are multivariate normal, the plot should form a relatively straight, diagonal line (see Figure 1). Points furthest from the line are observations that deviate most from the

Table 2. Means, Standard Deviations, and Correlations for Three Outcome Variables.

	College Major				
Variable	Humanities (n = 40), M (SD)	STEM (n = 40), M (SD)	1	2	3
1. MSC	3.75 (1.06)	4.88 (1.17)	—		
2. FOF	3.87 (1.13)	3.87 (1.42)	.51	—	
3. SC	3.56 (1.15)	3.64 (1.43)	.27	.28	—

Note. MSC = mathematics self-concept; FOF = fear of failure; SC = social competence.

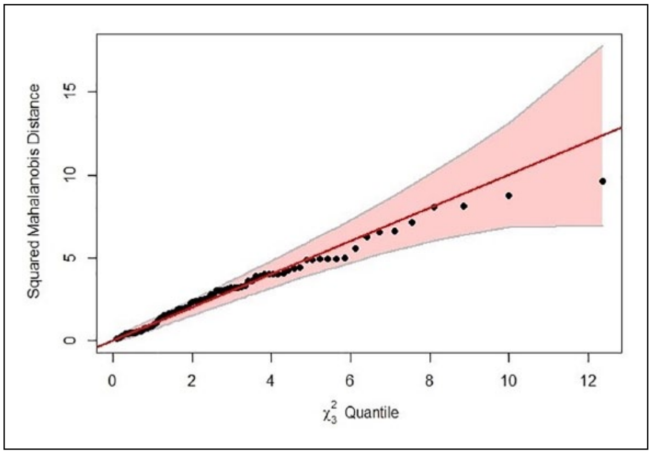

Figure 1. Chi-square Q-Q plot of Mahalanobis distances for two-group descriptive discriminant analysis model residuals, with 95% confidence envelope.

center of the multivariate distribution, indicating potential outliers. Importantly, an absence of univariate outliers does *not* suggest the same is true for the multivariate data, and although DDA is robust to violations of multivariate normality when sample sizes are large or group sizes are equal, results are particularly sensitive to multivariate outliers (Vallejo & Ato, 2012). For additional information, we refer readers to Leys et al. (2019) for discussion of different types of multivariate outliers and practical strategies for addressing each and to Leys, Klein, Dominicy, and Ley (2018) for a more technical treatment and discussion of robust D^2 estimation, along with SPSS and R syntax.

The homogeneity of the group variance/covariance matrices assumption is analogous to the homogeneity of variance assumption in ANOVA and is necessary to justify pooling the error variances across the groups (Nimon, 2012). Because the multivariate effect is expressed as a ratio of the between-group to within-group variance in the composite outcome, using the pooled variance/covariance matrix can bias the standard error estimate if the within-group matrices meaningfully differ. The most common way of evaluating the homogeneity assumption is Box's M test. Similar to Levene's test in ANOVA, Box's M tests the null hypothesis that the within-group variance/covariance matrices are equal to the pooled matrix. The test uses the log determinant of each matrix as a measure of generalized variance. Here, Box's M was *not* statistically significant (p = .06), suggesting the assumption was *not* violated.

However, Box's M is sensitive to even slight deviations from multivariate normality and may indicate that the assumption is violated (via a statistically significant result) when, in fact, the matrices are reasonably comparable across groups. The assumption can be further assessed through direct inspection of the log determinants for each group's covariance matrix, reported in the SSPS output before the Box's M test result. The smaller the ratio between these values, the more likely it is that the assumption holds, especially when group sizes are not radically different (Huberty & Olejnik, 2006; Tabachnick & Fidell, 2019). When group sizes do differ and the matrices are not homogeneous, violations are most problematic when group size and variance are negatively paired. For example, simulations of one- and two-way (Finch & French, 2013; Konietschke, Bathke, Harrar, & Pauly, 2015; Vallejo & Ato, 2012) designs have shown that when the largest and smallest cell sizes differed by a factor of two, log determinant ratios ≥10:1 in favor of the smallest group led to inflated Type 1 error rates, even when the outcomes were sampled from a multivariate normal distribution. Conversely, when the largest group had the greater generalized variance, Type 1 error rates were overly conservative and power to detect group differences was reduced. When the matrices were homogenous, differences between group sample sizes did not affect error rates.

In our example, the ratio of the two log determinants was 8:1. In conjunction with the nonsignificant Box's M and equal group sizes, we might conclude that the assumption is reasonably met. However, visualization is again strongly recommended to further evaluate the extent of the heterogeneity between the groups. Friendly and Sigal (2017, 2018) provide explanation and code for several visualizations of multivariate linear models. Figure 2 depicts one of these, a plot of the log determinants with surrounding asymptotic confidence intervals (see Cai, Liang, & Zhou, 2015) for each group matrix and the pooled covariance matrix. It shows that the confidence intervals for both groups overlap with each other and with the pooled determinant. This lends some additional support to the assumption, though the width of the confidence intervals reveals a large amount of uncertainty that should also be reported.

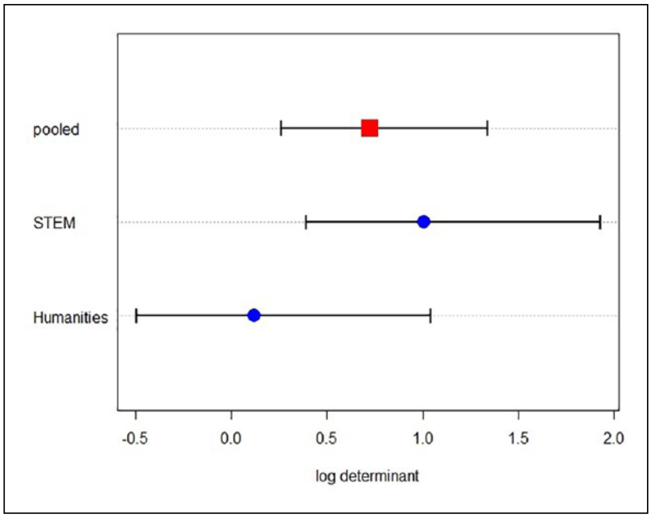

Figure 2. Log determinants of pooled and separate variance/covariance matrices with 95% confidence intervals for Humanities (n = 40) and STEM majors (n = 40).

Although we conclude that assumptions are satisfied for the purpose of this tutorial, researchers have several options when this is not the case. For instance, Vallejo and Ato (2012) found that a generalized least squares estimator was preferable to ordinary least squares in two-way heteroscedastic (ratios ≥12:1) MANOVAs under a variety of conditions, and Zhang, Zhou, Guo, and Liu (2016) introduced a modified Barlett test that performed better than classical tests for factorial MANOVAs with larger sample sizes. Most recently, Friedrich and Pauly (2018) developed a test statistic for MANOVA designs that does not require the distributional assumptions of multivariate normality or homogeneous matrices. Zimmermann, Pauly, and Bathke (2019) extended this approach to the multivariate analysis of covariance context, and Bathke et al. (2018) to repeated measures MANOVA. These options are implemented in the R package, MANOVA RM, which also includes a graphical user interface plugin to facilitate ease of use for researchers more familiar with SPSS.

Is There Anything Worth Interpreting?

Once assumptions are satisfied, the DDA results can be examined to determine whether there are any interpretable effects. In DDA, this means looking at the discriminant functions (also called *linear* or *canonical discriminant functions, dimensions*, or *roots*), which are the linear composites of the outcome variables that best separate the groups (Huberty & Olejnik, 2006). Each function represents a *unique* composite variable, derived from a *different* linear equation (Sherry, 2006). All functions are orthogonal, meaning no two functions use any of the same variance in the outcome variables to explain group differences. The first function creates a composite variable for which the group differences explain the greatest amount of variance possible, and each subsequent function creates a new composite variable to explain the greatest amount of the variance left over. For each hypothesis tested (i.e., for each main effect and interaction effect of interest), the number of functions generated in the analysis is equal to either one less than the number of groups ($k - 1$) or to the number of outcome variables, whichever is smallest (Enders, 2003). The researcher's task is to decide how many functions, if any, explain enough variance to warrant interpretation. Recalling the prior discussion, this decision should not be based solely on statistical significance but should always include effect size in the context of what is being studied and prior literature (Henson, 2006; Warne et al., 2012).

In the current example, there is only one function because there were two groups in the analysis. The function's value can be assessed by looking at its canonical correlation (R_c), included in the first box of the DDA output under the heading "Summary of Canonical Discriminant Functions." The R_c is simply the correlation between the composite variable (created from the observed outcome variables) and the grouping variable. Squared, the R_c is a variance-accounted-for effect size (R_c^2) that is analogous to R^2 in multiple regression or η^2 in ANOVA (Sherry, 2006). In our example, the function had an R_c of .56 and an R_c^2 of .31, indicating the grouping variable could account for 31% of the variance in the composite outcome variable. This effect size can also be calculated using Wilks's lambda (Λ or λ), reported in the next box of the output. This statistic has the beneficial property of being a reverse variance-accounted-for effect size, such that $1 - \Lambda$ is equal to the total variance explained by group membership (Enders, 2003). For instance, in our example $\Lambda = .72$, and $1 - \Lambda = .28$, which in turn is equal to the squared canonical correlation.

Wilks's lambda is also the statistic used to test for statistical significance. The number of statistical significance tests included in the output will equal the number of functions. However, these tests occur in hierarchical fashion, with the first testing the full model, the second testing all subsequent functions, and so on. Only the last function is tested in isolation (Sherry, 2006). With only two groups in the analysis, the single function is equivalent to the full model, which in this case was statistically significant ($\Lambda = .72$, $\chi[3] = 25.65$, $p < .001$). Because the function also explained a substantial portion of the between group variance, we can conclude there is indeed an effect worth interpreting.

Where Is the Effect Coming From?

Having identified a considerable difference between the two groups, it is reasonable to investigate its origin and determine how each outcome variable contributed. As noted, running three univariate ANOVAs as post hocs to test mean differences on each outcome independently can be a misleading strategy. Rather, researchers can consult the standardized

Table 3. Standardized Discriminant Function Coefficients, Structure Coefficients, and Group Centroids.

Outcome variable	Standardized coefficient	r_s	r_s^2
MSC	1.23	.86	.74
FOF	−0.67	<.01	<.01
SC	−0.13	.06	<.01

Group	Centroids [95% CI]	Cohen's *d* [95% CI]
Humanities	−0.62 [−0.94, −0.31]	1.25 [0.76, 1.72]
STEM	0.62 [0.31, 0.94]	

Note. MSC = mathematics self-concept; FOF = fear of failure; SC = social competence; r_s = structure coefficient; CI = confidence interval; $d = \bar{X}_1 - \bar{X}_2 / SD_{pooled}$.

weights and structure coefficients provided by the DDA output to examine the makeup of the multivariate composite variable itself as well as the group centroids to determine the direction and magnitude of the difference. These results are presented in Table 3.

In DDA, the weights are called standardized discriminant function coefficients, and they tell us the contribution of each outcome variable to the linear equation that created the composite outcome variable for the function. Looking at Table 3, mathematics self-concept made the greatest contribution to the equation with a standardized function coefficient of 1.23, followed by fear of failure with a standardized function coefficient of −0.67. Social competence played virtually no role in generating the composite outcome variable scores. Based *solely* on these coefficients, we would conclude mathematics self-concept and fear of failure best discriminated between the groups.

However, this inference is premature because standardized weights cannot be taken as measures of variable importance in GLM analyses. Doing so leads to at best, incomplete, and at worst, inaccurate interpretative conclusions (Thompson, 2006), particularly in two situations. The first is when the variables in the analysis are multicollinear. The presence of multicollinearity means more than one variable can explain at least a portion of the same variance in the effect. Because variables cannot receive credit for the same variance in the linear equation, shared variance is either divided in some fashion among the variables or attributed entirely to one. Of course, these divisions of credit are formulaically determined but slight variation in the degree of multicollinearity can have substantial impact on the nature of the division, making the process appear somewhat arbitrary. The standardized weights thus only indicate the contribution each variable actually made to the equation and not the contribution each *could* have made (Henson, 2002). As a consequence, a researcher might dismiss a variable with a low weight as unimportant, despite its ability to explain a substantial amount of the variance.

The reverse situation is also problematic, in which a variable can explain hardly any of the effect, but nonetheless receives a substantial amount of credit in the linear equation. This is known as a suppression effect (Conger, 1974). A suppressor variable contributes to the overall prediction by strengthening the relationship between the composite and one or more of the other variables in the analysis, despite having little or no relationship with the composite itself (Warne, 2014). Interpretations based solely on weights cannot detect suppression effects, leading the researcher to mistakenly identify a suppressor variable as having an important relationship with the composite while missing its actual role in the effect.

The interpretive problems resulting from multicollinearity and suppression are brought to light when structure coefficients are consulted in conjunction with standardized weights to identify variable relationships in GLM analyses (Courville & Thompson, 2001). A structure coefficient (r_s) is the bivariate correlation between an observed variable and a composite variable. In a squared metric (r_s^2), structure coefficients indicate the amount of variance a given variable can explain in the composite, irrespective of other variables in the analysis (Henson, 2002). Accordingly, whereas the weights denote each variable's contribution to the regression equation, structure coefficients identify the actual relationship between the observed variables and the composite. A variable with a low weight and high structure coefficient suggests the variance the variable could have explained is being credited to another variable as a result of multicollinearity. Conversely, a variable with a high weight and low structure coefficient suggests it is acting as a suppressor in the analysis (Kraha, Turner, Nimon, Zientek, & Henson, 2012). These meaningful interpretations cannot be determined with only examination of standardized weights.

Returning to our example, the structure coefficients in Table 3 reveal that fear of failure has essentially no relationship with the composite outcome variable, explaining less than 1% of its variance. Nevertheless, its standardized function coefficient indicates it did contribute to the linear equation used to create the composite outcome variable. In fact, if the DDA was run without fear of failure in the analysis, the R_c^2 decreases to .22, a full 8% difference in the effect size. This drop is substantially more than would be anticipated by the squared structure coefficient. It is acting as a suppressor, helping mathematics self-concept do a better job of explaining variance in the composite outcome despite not being able to explain much itself. Relying on the function coefficients alone would have caused us to miss this suppression effect and assume fear of failure played a direct role in separating the groups. Although it might be unclear in many cases why suppression is occurring, such a result can inform future questions about the nature of the variable relationships, including how suppressors should be modeled in subsequent research (Kim, 2019).

Once the makeup of the composite outcome variable is understood, the final step is to examine the group centroids, the last piece of relevant information in the DDA output (the subsequent boxes under “Classification Statistics” in the SPSS output refer to PDA results). The group centroids are the means of each group on the composite outcome variable that was created from the observed variables, and they inform the researcher which groups are different and in what direction (Huberty & Olejnik, 2006). With only two groups in our example, we know which groups are different, but the centroids indicate it was the STEM group that had higher average scores on the composite outcome. Fear of failure’s role as a suppressor should also be noted when reporting the results.

To determine the magnitude of the group difference, researchers using SPSS can run an ANOVA with the newly saved multivariate function scores as the dependent variable. Lest the reader think we have forgotten our own admonitions against univariate tests, we emphasize that using the composite variable in the analysis ensures the results ultimately remain multivariate in nature. The ANOVA results provide the confidence intervals shown in Table 3 as well as the standard deviations for each group that we can use to calculate a Cohen’s d effect size. Here, $d = 1.25$, 95% CI [0.76, 1.72], a large effect suggesting a substantial difference in the composite outcome between the two groups.

DDA as Post Hoc to Factorial MANOVA

DDA can also be used as a post hoc analysis to interpret interaction effects in a factorial MANOVA. To demonstrate this procedure, the preceding illustration was expanded to include a four-level grouping variable based on mathematics and verbal ability profiles. Groups include students with profiles of high mathematics and high verbal ability ($n = 20$), high mathematics but low verbal ability ($n = 20$), low mathematics and low verbal ability ($n = 20$), and high verbal but low mathematics ability ($n = 20$). The research question is whether there is an interaction effect between ability profile and college major (humanities, $n = 40$; STEM, $n = 40$) on the same three outcome variables (mathematics self-concept, fear of failure, and social competence), resulting in a 4×2 balanced MANOVA. We use the same strategy to frame the interpretation, highlighting differences from the two-group scenario.

SPSS Procedures. Although no point and click method is available in SPSS to run a DDA of an interaction effect, DDA results can be readily obtained by simply adding a “DISCRIM” subcommand to the syntax used to run a factorial MANOVA. This syntax will produce the traditional omnibus MANOVA results, along with separate DDA results for the interaction effect and each main effect, though only the interaction effect is interpreted here. For either of the main effects, the same procedure discussed above could be used.

The focus of this analysis is the multivariate linear combination of the observed outcome variables that yields the strongest possible interaction effect. It is important to emphasize that each of the main effects and the interaction effect will have its own composite outcome variable, and therefore, each can be meaningfully different in interpretation. This is a difference between univariate and multivariate approaches, which points again to the need to interpret a MANOVA interaction effect with multivariate methods.

In SPSS, the discriminant function scores for an interaction effect cannot be automatically saved in the data set as done above (though they can be in R). However, they can be easily computed in two steps. First, each outcome variable must be standardized. This is done by subtracting the variable mean from the observed scores and dividing by the square root of the mean square error term (Enders, 2003). The mean square error for each outcome variable can be found in the univariate ANOVA results, which are included in the MANOVA output. In our example, the syntax to compute the standardized outcome variables is

$$\text{COMPUTE zMSC} = (\text{MSC} - 4.32)/\text{sqrt}(1.14)$$
$$\text{COMPUTE zFOF} = (\text{FOF} - 3.87)/\text{sqrt}(1.31)$$
$$\text{COMPUTE zSC} = (\text{SC} - 3.6)/\text{sqrt}(1.24)$$

Second, the interaction composite variable is computed via the linear equation, which multiplies the standardized outcome variables with their respective standardized function coefficients (weights), which are listed in the MANOVA output as a result of the “DISCRIM” subcommand (see also Table 4).

$$\text{COMPUTE int\_comp} = (-0.45 * \text{zMSC}) + (0.63 * \text{zFOF}) + (0.89 * \text{zSC}).$$

If more than one function is interpretable, the second step can be repeated using the second function’s weights.

Descriptives and Assumptions

The means, standard deviations, and correlations for each outcome are reported in Table 5. The chi-square Q-Q plot revealed no major issues with multivariate normality or multivariate outliers (Figure 3). Box’s M test was again statistically nonsignificant ($p = .68$), and the ratio between the

Table 4. Standardized Discriminant Function Coefficients, Structure Coefficients, and Group Centroids for Major by Ability Interaction Effect on Three Outcomes.

Outcome variable	Standardized coefficient	r_s	r_s^2
MSC	−0.45	.11	.01
FOF	0.63	.48	.23
SC	0.89	.85	.72

Cell	Centroids [95% CI]	Cohen's *d* [95% CI]
HMHV		
Humanities	0.91 [0.28, 1.54]	1.18 [0.21, 2.12]
STEM	−0.33 [−0.96, 0.30]	
HMLV		
Humanities	0.30 [−0.33, 0.93]	1.41 [0.41, 2.38]
STEM	−1.07 [−1.70, −0.44]	
LMHV		
Humanities	−0.17 [−0.80, 0.46]	−1.81 [−2.84, −0.73]
STEM	1.43 [0.80, 2.06]	
LMLV		
Humanities	−0.21 [−0.84, 0.43]	0.61 [−0.30, 1.50]
STEM	−0.87 [−1.50, −0.24]	

Note. MSC = mathematics self-concept; FOF = fear of failure; SC = social competence; HMHV = high mathematics/high verbal; HMLV = high mathematics/low verbal; LMHV = low mathematics/high verbal; LMLV = low mathematics/low verbal; r_s = structure coefficient; $d = \bar{X}_1 - \bar{X}_2 / SD_{pooled}$; CI = confidence interval.

Table 5. Means, Standard Deviations, and Correlations for Three Outcome Variables.

	Outcome variable		
Cell (*n* = 10)	MSC, *M* (*SD*)	FOF, *M* (*SD*)	SC, *M* (*SD*)
HMHV			
Humanities	3.99 (0.80)	3.83 (1.12)	4.59 (0.99)
STEM	5.57 (1.22)	4.27 (1.33)	3.59 (1.40)
HMLV			
Humanities	3.96 (0.81)	4.36 (0.88)	3.46 (0.72)
STEM	4.85 (1.13)	3.34 (0.81)	2.91 (1.46)
LMHV			
Humanities	3.58 (1.41)	3.87 (1.10)	3.00 (1.06)
STEM	5.06 (0.66)	5.05 (1.04)	4.98 (0.73)
LMLV			
Humanities	3.48 (1.15)	3.44 (1.26)	3.20 (1.18)
STEM	4.04 (1.15)	2.81 (1.27)	3.10 (1.14)
1	—		
2	.51	—	
3	.27	.28	—

Note. MSC = mathematics self-concept; FOF = fear of failure; SC = social competence; HMHV = high mathematics/high verbal; HMLV = high mathematics/low verbal; LMHV = low mathematics/high verbal; LMLV = low mathematics/low verbal.

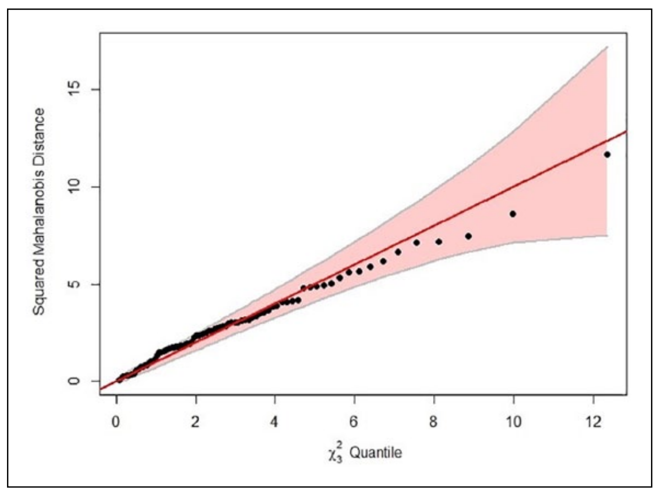

Figure 3. Chi-square Q-Q plot of Mahalanobis distances of factorial multivariate analysis of variance model residuals.

largest and smallest log determinants was 3:1, providing initial evidence that the variance/covariance matrices for each cell were reasonably similar. Graphical examination of the log determinants supported this result (Figure 4), this time indicating a more precise estimate of the pooled variance than in the two-group example.

Is There Anything Worth Interpreting?

The first part of the interaction output displays four multivariate statistical significance tests of the full model: Hotelling–Lawley trace, Pillai–Bartlett trace, Wilks's lambda, and Roy's greatest characteristic root (only Wilks's lambda is included in the SPSS output when a DDA is run separately).

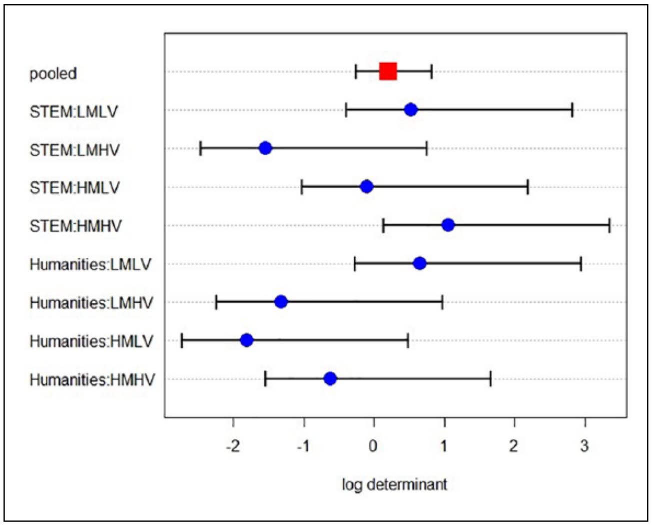

Figure 4. Log Determinants of pooled and separate variance/covariance matrices with 95% confidence intervals for college major ($k = 2$) by ability profile ($k = 4$) interaction cells ($n = 10$ for all groups).
Note. HMHV = high mathematics/high verbal; HMLV = high mathematics/low verbal; LMHV = low mathematics/high verbal; LMLV = low mathematics/low verbal.

All four will be identical when only two groups are in the analysis and will typically produce consistent results in the multiple group case, especially when group sample sizes are similar. When the outcome variables are highly correlated and suggest a single underlying dimension (a concentrated eigenvalue pattern), Roy's test has the greatest power and Pillai–Bartlett the least. With multiple dimensions (a diffuse eigenvalue pattern), Pillai–Bartlett is most powerful, followed by Wilks's lambda, Hotelling–Lawley, and then Roy's (Huberty & Olejnik, 2006). Wilks's lambda is most often reported because it tends to lie in the middle of these options and as mentioned above, it expresses the unexplained variance in the model. Researchers should always specify which test results they are reporting and explain if the results were inconsistent. In this example, using the Wilks's lambda criterion, the full model of all possible interaction effect functions was statistically significant, $\Lambda = .65$, $F(9, 171) = 3.63$, $p < .001$, and able to explain 35% of the variance in the composite outcome variables ($1 - \Lambda = .35 = R_c^2$).

At this point, the interpretation diverges somewhat from the prior two-group situation. Recall that the number of functions generated in a DDA is equal to the smaller of either the number of outcomes or $k - 1$ (Enders, 2003). Here, the smaller value is the number of outcome variables, so three functions are included in our output. The 35% variance explained by the full model noted above accounts for all of them, even though all functions might not individually explain enough variance to be noteworthy. Function 1 will explain the largest possible portion of variance, Function 2 will explain the greatest amount possible *out of what is left over*, and Function 3 will explain as much as possible *out of the remainder*. Because each function represents a different composite variable, they must be considered separately to decide how many are worth interpreting.

By squaring the canonical correlations for each function, we find that Function 1 accounted for 28% of the total effect and Function 2 explained 8%. The test of Functions 2 and 3 together was not statistically significant ($p = .14$). In practice, the researcher would need to determine if the 8% of variance explained by Function 2 is worth interpreting, though we only interpret Function 1 in this example.

Where Is the Effect Coming From?

We again examine the standardized function coefficients and structure coefficients of each outcome variable to unpack the interaction composite in multivariate fashion. These coefficients and centroids for Function 1 are reported in Table 4. Although the output includes this information for each function, because we have decided only Function 1 is interpretable, the results for Functions 2 and 3 are not reported. In practice, researchers should report the results for any function they interpret.

The results indicate social competence received the most credit in the linear equation and was also able to explain the greatest amount of variance in the interaction composite variable ($r_s^2 = .72$). Fear of failure made a smaller but still sizable contribution to the effect and was able to account for 21% of the composite variance. The discrepancy between the function and structure coefficients for mathematics self-concept suggests it may have acted as a small suppressor in the analysis. This interpretation of the composite outcome variable tells us that the interaction effect observed is mainly on social competence, but with secondary contribution from fear of failure.

With more than two groups in the analysis, we now need to determine which groups are different on the interaction composite. The simplest way to do this is to use the multivariate interaction composite we created above with the Function 1 weights as the dependent variable in a 4×2 ANOVA. Again, we emphasize here that use of the composite ensures the analysis remains multivariate. To aid interpretation, it is helpful to display the ANOVA results in an interaction graph depicting each group's centroid or mean function score.

The interaction graph (Figure 5) indicates that, on average, humanities majors with profiles characterized by overall high ability as well as those with high mathematics but low verbal ability had higher function scores than STEM majors of similar ability profiles. However, the reverse was true for students of low mathematics but high verbal ability, with STEM majors with this ability profile having the highest mean function score of any group. The average function score for STEM and humanities majors with low ability in both areas did not differ statistically. Calculated Cohen's *d*

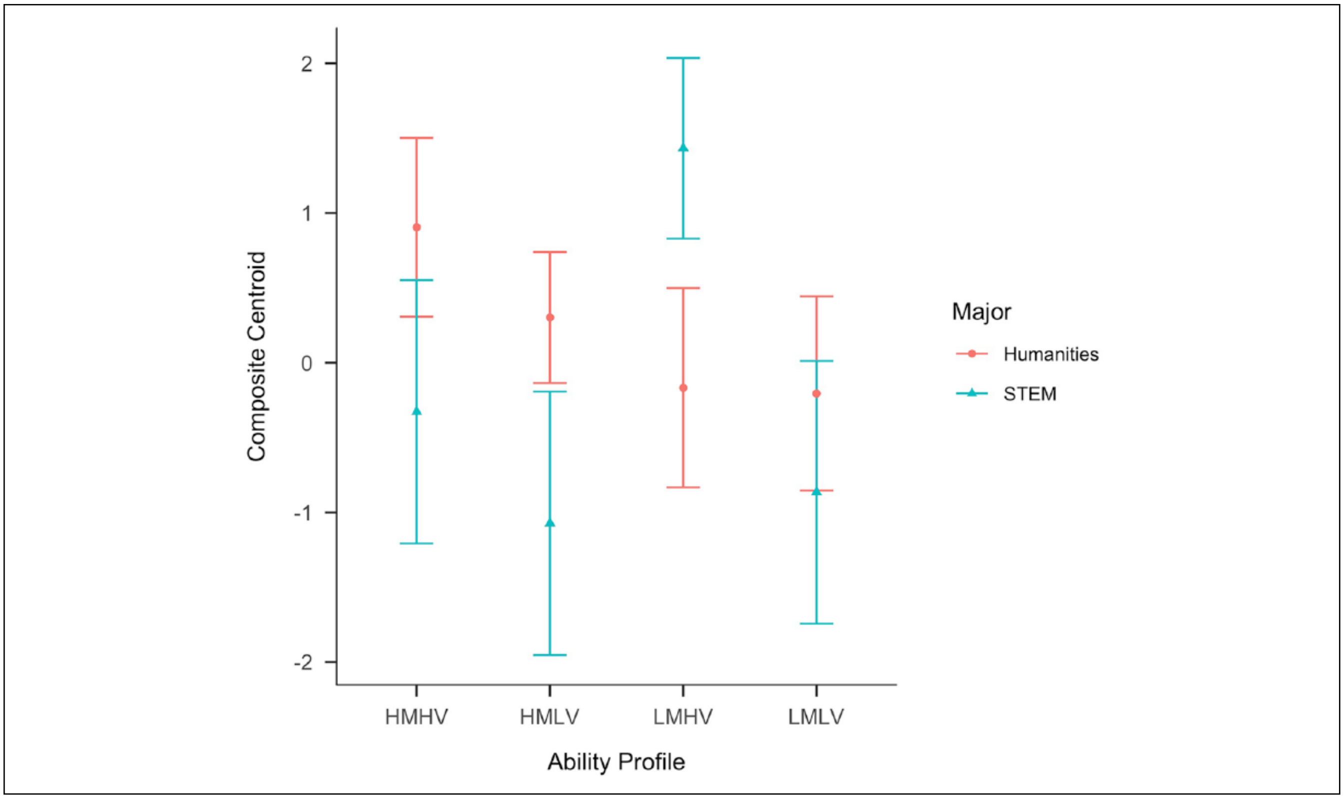

Figure 5. Major by ability profile interaction plot of composite centroids with 95% confidence intervals (n = 10 for all groups).
Note. MSC = mathematics self-concept; FOF = fear of failure; SC = social competence; HMHV = high mathematics/high verbal; HMLV = high mathematics/low verbal; LMHV = low mathematics/high verbal; LMLV = low mathematics/low verbal.

effect sizes revealed differences between STEM and humanities majors were strongest for students with low mathematics and high verbal ability in favor of STEM majors ($d = -1.81$, 95% CI [−2.84, −0.73]) followed by those with high mathematics but low verbal ability in favor of humanities majors (d = 1.41, 95% CI [0.41, 2.38]). Because the function was composed primarily of social competence, with a secondary contribution from fear of failure, these results suggest students in our sample with low ability in the domain most associated with their major but high ability in the opposite domain, had higher social competence, and were slightly more afraid of failing than students with the same ability profile but opposite major.

Conclusion

This article has attempted to bring conceptual and procedural clarity to using MANOVA to describe group differences on related outcome variables in gifted education research. Given their greater ability to model real-world complexity, multivariate approaches like MANOVA and DDA are often preferable to their univariate counterparts. However, the advantages of multivariate methods are only realized when accompanied by multivariate interpretations. We hope that the explanation and demonstration provided here will alert researchers to the problems associated with univariate post hoc testing in multivariate contexts as well as to the alternative benefits of DDA as a MANOVA post hoc option. Though we have tried to capture the fundamental decisions and reporting requirements typical of these analyses, we emphasize the limits of our example. No heuristic (or default software setting) can supplant the need for reflective researcher judgment. Analytic decisions and statistical interpretations should always be justified in relation to a particular research purpose and study design.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Kendal N. Smith https://orcid.org/0000-0002-2266-7901
Kristen N. Lamb https://orcid.org/0000-0003-0488-8863

References

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*, 3-25. doi:10.1037/amp0000191

Barton, M., Yeatts, P. E., Henson, R. K., & Martin, S. B. (2016). Moving beyond univariate post-hoc testing in exercise science: A primer on descriptive discriminant analysis. *Research Quarterly for Exercise and Sport*, *87*, 365-375. doi:1080/02701367.2016.1213352

Bathke, A. C., Friedrich, S., Pauly, M., Konietschke, F., Staffen, W., Strobl, N., & Höller, Y. (2018). Testing mean differences among groups: Multivariate and repeated measures analysis with minimal assumptions. *Multivariate Behavioral Research*, *53*, 348-359. doi:10.1080/00273171.2018.1446320

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological*, *57*, 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Boedeker, P., & Kearns, N. T. (2019). Linear discriminant analysis for prediction of group membership: A user-friendly primer. *Advances in Methods and Practices in Psychological Science*, *2*, 250-263.

Cai, T. T., Liang, T., & Zhou, H. H. (2015). Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. *Journal of Multivariate Analysis*, *137*, 161-172. doi:10.1016/j.jmva.2015.02.003

Chae, Y., & Gentry, M. (2011). Gifted and general high school students' perceptions of learning and motivational constructs in Korea and the United States. *High Ability Studies*, *22*, 103-118. doi:10.1080/13598139.2011.577275

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426-443. doi:10.1037/h0026714

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, *34*, 35-46.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: B is not enough. *Educational and Psychological Measurement*, *61*, 229-248. doi: 10.1177/0013164401612006

Enders, C. K. (2003). Performing multivariate group comparisons following a statistically significant MANOVA. *Measurement and Evaluation in Counseling and Development*, *36*, 40-56.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Thousand Oaks, CA: Sage.

Finch, H., & French, B. (2013). A Monte Carlo comparison of robust MANOVA test statistics. *Journal of Modern Applied Statistical Methods*, *12*, 35-81. doi:10.22237/jmasm/1383278580

Fish, L. J. (1988). Why multivariate methods are usually vital. *Measurement and Evaluation in Counseling and Development*, *21*, 130-137. doi:10.1080/07481756.1988.12022895

Friedrich, S., & Pauly, M. (2018). MATS: Inference for potentially singular and heteroscedastic MANOVA. *Journal of Multivariate Analysis*, *165*, 166-179. doi:10.1016/j.jmva.2017.12.008

Friendly, M., & Sigal, M. (2017). Graphical methods for multivariate linear models in psychological research: An R tutorial. *Quantitative Methods for Psychology*, *13*, 20-45. doi:10.20982/tqmp.13.1.p020

Friendly, M., & Sigal, M. (2018). Visualizing tests for equality of covariance matrices. *The American Statistician*. Advance online publication. doi:10.1080/00031305.2018.1497537

Graham, J. M. (2008). The general linear model as structural equation modeling. *Journal of Educational and Behavioral Statistics*, *33*, 485-506.

Grice, J. W., & Iwasaki, M. (2007). A truly multivariate approach to MANOVA. *Applied Multivariate Research*, *1*, 199-226. Retrieved from https://amr.uwindsor.ca/ojs/leddy/index.php/AMR

Henson, R. K. (2002, April). *The logic and interpretation of structure coefficients in multivariate general linear model analyses*. Paper presented at the annual meeting of the American Educational Research Association, *New Orleans*. (ERIC Document Reproduction Service No. ED 467381)

Henson, R. K. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist*, *34*, 601-629. doi:10.1177/0011000005283558

Huang, F. L. (2018). Multilevel modeling and ordinary least squares regression: How comparable are they? *Journal of Experimental Education*, *86*, 265-281. doi:10.1080/00220973.2016.1277339

Huang, F. L. (2020). MANOVA: A procedure whose time has passed? *Gifted Child Quarterly, 64*, 56-60.

Huberty, C. J., & Hussein, M. H. (2003). Some problems in reporting use of discriminant analyses. *Journal of Experimental Education*, *71*, 177-192. doi:10.1080/00220970309602062

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, *105*, 302-308. doi:10.1037/0033-2909.105.2.302

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis*. Hoboken, NJ: Wiley.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., . . .Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, *68*, 350-386. Retrieved from http://www.jstor.org/stable/1170601

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in AERJ and JCP articles from 1988 to 1997: A methodological review. *Journal of Experimental Education*, *69*, 280-309. doi:10.1080/00220970109599489

Kim, Y. (2019). The causal structure of suppressor variables. *Journal of Educational and Behavioral Statistics*, *44*, 367-389. doi:10.3102/1076998619825679

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin*, *85*, 410-416. doi:10.1037/0033-2909.85.2.410

Konietschke, F., Bathke, A. C., Harrar, S. W., & Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis*, *140*, 291-301. doi:10.1016/j.jmva.2015.05.001

Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology*, *3*, 44.

Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3303138/

Leys, C., Delacre, M., Mora, Y. L., Läkens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, *32*(1), 5. doi:10.5334/irsp.289

Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology*, *74*, 150-156. doi:10.1016/j.jesp.2017.09.011

Maxwell, S. (1992). *Recent developments in MANOVA applications*. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 2, pp. 137-168). Greenwich, CT: JAI Press.

McCoach, D. B. (2010). Dealing with dependence (Part II): A gentle introduction to hierarchical linear modeling. *Gifted Child Quarterly*, *54*, 252-256. doi:10.1177/0016986210373475

McCoach, D. B., & Adelson, J. L. (2010). Dealing with dependence (Part I): Understanding the effects of clustered data. *Gifted Child Quarterly*, *54*, 152-155. doi:10.1177/0016986210363076

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*, 114-140. doi:10.1037/met0000078

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195-244. doi:10.2466/pr0.1990.66.1.195

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini-review. *Frontiers in Psychology*, *3*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3428812/

Nimon, K. F. (2015). Secondary data analyses from published descriptive statistics: Implications for human resource development theory, research, and practice. *Advances in Developing Human Resources*, *17*, 26-39. doi:10.1177/1523422314559805

Pyryt, M. C. (2004). Pegnato revisited: Using discriminant analysis to identify gifted children. *Psychology Science*, *46*(3), 342-347.

Puryear, J. S., & Kettler, T. (2017). Rural gifted education and the effect of proximity. *Gifted Child Quarterly*, *61*, 143-152. doi:10.1177/0016986217690229

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Sherry, A. (2006). Discriminant analysis in counseling psychology research. *The Counseling Psychologist*, *34*, 661-683. doi:101177/0011000006287103

Sherry, A., & Henson, R. K. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment*, *84*, 37-48. doi:10.1207/s15327752jpa8401_09

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366. doi:10.1177/0956797611417632

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Boston, MA: Pearson.

Thompson, B. (1999, April). Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. Paper presented at the annual meeting of the American Educational Research Association, *Montreal, Canada*. (ERIC Document Reproduction Service No. ED 429 110)

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York, NY: Guilford Press.

Thompson, B. (2015). The case for using the general linear model as a unifying conceptual framework for teaching statistics and psychometric theory. *Journal of Methods and Measurement in the Social Sciences*, *6*(2), 30-41. doi:10.2458/v6i2.18801

Tonidandel, S., & LeBreton, J. M. (2013). Beyond step-down analysis: A new test for decomposing the importance of dependent variables in MANOVA. *Journal of Applied Psychology*, *98*, 469-477. doi:10.1037/a0032001

Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists. *Practical Assessment, Research & Evaluation*, *19*(17), 1-10.

Warne, R. T., Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly*, *56*, 134-149. doi:10.1177/0016986212444122

Vallejo, G., & Ato, M. (2012). Robust tests for multivariate factorial designs under heteroscedasticity. *Behavior Research Methods*, *44*, 471-489. doi:10.22606/jas.2016.12006

Zhang, J. T., Zhou, B., Guo, J., & Liu, X. (2016). A modified Bartlett test for heteroscedastic two-way MANOVA. *Journal of Advanced Statistics*, *1*(2), 94-108. doi:10.22606/jas.2016.12006

Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher*, *38*, 343-352. doi:10.3102/0013189X09339056

Zimmermann, G., Pauly, M., & Bathke, A. C. (2019). *Multivariate analysis of covariance when standard assumptions are violated*. Retrieved from https://arxiv.org/abs/1902.10195v1

Author Biographies

Kendal N. Smith is a doctoral candidate in the Department of Educational Psychology at the University of North Texas. Her work considers research methods in psychology and education as well as intersections between creativity, morality, and epistemic cognition.

Kristen N. Lamb is a research associate for the Robinson Center at the University of Washington. Her interests include equity issues in gifted education, impacts of educational policy on teacher practice and student achievement, the role of creativity in talent development of high-ability students, and classroom conditions conducive to developing creative thinking and advanced academic achievement.

Robin K. Henson is a professor and Chair in the Department of Educational Psychology and holds an appointment as a University Distinguished Teaching Professor at the University of North Texas. He teaches in the research methodology, measurement, and statistics area, and his primary research areas include applied statistical analysis, graduate-level methodology training, and self-efficacy theory.

Manuscript received: June 28, 2018; Final revision received: August 15, 2019; Accepted: October 17, 2019