MANOVA Post Hoc Techniques Used in
Published Articles: A Systematic Review

Fatimah A. Al-Abdullatif   **Mohammed A. Al-Abdullatif**   **Gordon Brooks**
Ohio University

Researchers are usually interested in examining group mean differences after exhibiting significant MANOVA test. The MANOVA-ANOVA approach has been recognized as an issue in several previous research reviews. The enormous growth of social and behavioral research in the past decade has necessitated the need for and use of research reviews. Therefore, the purpose of this review is to identify trends in data-analytic practices in pertain to MANOVA follow-up procedures and to examine what specific tests have been employed. Articles from five major databases (e.g. ERIC, PsychINFO) that regularly publish research in education were examined. The results of analyzing 235 studies were reported. The results indicated that to explain significant MANOVA results, univariate tests were the most post hoc method applied in 95% of the studies, whereas multivariate methods such as DFA was only applied in 5% of the studies.

Research synthesis is one of the most important components of conducting research, as researchers acquire a better understanding of the topic and become cognizant of the field. Within the body of literature, there is a wealth of statistical methods that can be explored to identify methodological trends and research movements characterizing the field (Kieffer, Reese, & Thompson, 2001). The enormous growth of social and behavioral research in the past decade has necessitated the need for and use of research reviews (Cooper, Hedges, & Valentine, 2009). Therefore, it is imperative that researchers review the current state of the field and evaluate its content to better direct future efforts (Warne, Lazo, Ramos, & Ritter 2012).

The MANOVA-ANOVA approach has been an issue in several previous research reviews (cf. Keselman et al., 1998; Kieffer et al., 2001). Huberty and Morris (1989) found that when particular cases of univariate data analyses were comprised of several dependent variables and when multivariate analysis was conducted as the main analysis, following up with univariate analysis was not appropriate to exploring the identified multivariate effects (Kieffer et al., 2001). Borgen and Seling (1978) contended that once MANOVA is applied, which inherently suggests that the data are actually multivariate, a multivariate follow-up procedure is needed to discern the intricacy of the data that DFA (a multivariate technique) can do and univariate ANOVA cannot. Thompson (1999) found that when researchers follow up using univariate tests, they tend to overlook the information regarding the differences in the variate analyzed in the multivariate analysis. Kieffer et al. (2001) opined that "It is illogical first to declare interest in a multivariate system of variables and then to explore detected multivariate effects by conducting non-multivariate tests" (p. 287).

Purpose

Reviewing methodological procedures used in the literature is an important practice for many reasons (Warne et al., 2012). Most importantly, such reviews aid researchers in knowing what methodological learning is appropriate to understanding the quantitative investigation conducted in respective fields. Furthermore, research reviews delineate methodological trends that occurred throughout earlier years. Besides, reviewing statistical techniques allows researchers to assess the soundness of their field in terms of the recognized standards of statistical reporting (Warne et al., 2012). The purpose of this review is to identify trends in data-analytic practices that pertain to MANOVA follow-up procedures and to look at what specific tests have been employed.

We first summarize certain findings from previous research reviews concerning the MANOVA-ANOVA approach. Second, we present a context for describing follow-up practices. Third, we illustrate five years of follow-up practices from the top five databases, representing 177 journals of different fields, and evaluate them within the focus of our topic. We end with a discussion on the statistical status in following MANOVA in social research. Therefore, it is the primary intention of this review to look at five main areas: (a) the type of post hoc applied; (b) the most frequently used multivariate test; (c) the most frequently used univariate post hoc; (d) how frequently multivariate or univariate homogeneity tests, such as Levene's test and Box's M, are reported; and (e) the type of software packages used.

Previous Research Reviews

Over the past years, many scholars have explored and discussed the occurrence of univariate post hoc methods for significant MANOVA. They tackled the topic as a statistical methods issue used in different educational and psychological journals (cf. Armstrong & Henson, 2005; Barton, Henson, & Martin, 2016; Huberty & Morris, 1989; Keselman et al., 1998; Kieffer et al., 2001; Tonidandel & LeBreton, 2013; Warne et al., 2012).

In their review of five APA research journals published from 1948 to 1972, namely, the *Journal of Applied Psychology, Journal of Counseling Psychology, Journal of Consulting and Clinical Psychology (JCP), Developmental Psychology,* and the *Journal of Educational Psychology*, as well as one AERA journal, the *American Educational Research Journal (AERJ),* Huberty and Morris (1989) found that most studies using MANOVA carried out follow-ups with multiple ANOVAs. Out of 88 analyses of MANOVA, only two studies considered DA alongside multiple ANOVAs to be the dominant interpretation focus. Keselman and his colleagues (1998) found comparable results from reviewing articles in 17 leading journals published between 1994 and 1995. Out of a total of 79 identified articles, 66 (84%) studies employed multiple univariate analyses. From those 79 reviewed studies, DA procedures were utilized by four studies only.

Interestingly, in Kieffer et al.'s (2001) methodological review of approximately 10 years (1988 to 1997) of published work in two influential research journals, *AERJ* and *JCP,* univariate techniques as post hoc methods for examining multivariate effects were applied with high frequency: 21 out of 29 studies (72%) in AERJ, and 124 out of 160 studies (77%) in JCP. Univariate post hoc analyses, such as the Scheffé and Tukey tests, were generally reported in both journals, whereas DA was reported infrequently.

A more recent review by Armstrong and Henson (2005) examined 37 research studies in the *International Journal of Play Therapy* between 1993 and 2003. They found that two out of only three studies using MANOVA employed univariate ANOVAs to follow-up on significant multivariate results. Warne et al. (2012) examined articles published between 2006 and 2010 from five major journals of gifted education: The *Gifted Child Quarterly, High Ability Studies*, *Journal of Advanced Academics, Journal for the Education of the Gifted,* and the *Roeper Review*. Sixty-two articles were identified for employing MANOVA/MANCOVA. Half of all the articles (n = 31) applied a series of ANOVAs to follow-up on significant effects. Only five (8%) of those studies applied DA and only one used the Roy-Bargmann test as post hoc methods. Moreover, in an examination of 58 MANOVA articles published in three major psychology journals *(JCP*, *Emotion*, and *the Journal of Counseling Psychology)* between 2009 and 2013, Warne (2014) found no single study to have used DA. Analogous to Warne's (2014) findings, Tonidandel and LeBreton (2013) examined articles published in the *Journal of Applied Psychology* over a three-year period and found that most studies using MANOVA did not apply DA but instead relied on multiple ANOVAs to explore significant effects.

Bird and Hadzi-Pavlovic (2014) conducted a PsycARTICLES search of research work published between 2001 and 2010 using the keyword "manova." Out of 1,128 articles, 100 studies involving significant MANOVA results were examined. Ninety-six of these studies were further assessed by univariate post hoc on single variates; DA was considered in only three of these 96 articles, accompanied by univariate follow-up tests. For the remaining four studies, two involved DA, one article included the Roy-Bargmann step-down analysis, and one study specified only patterns of interpretations of mean differences on single outcome variables.

In their content analysis of four sport and exercise psychology journals (the *RQES, Psychology of Sport and Exercise, Journal of Sport and Exercise Psychology,* and the *Journal of Physical Activity and Health*), Barton et al. (2016) examined how often univariate follow-up tests, such as ANOVA, Tukey's HSD test, and Bonferroni correction, are utilized in lieu of a multivariate method DFA. A total of 81 articles, in the 2014 year, were identified as having examined group differences on multiple outcome variables. Of these studies, 30 used MANOVA and only one considered DA, whereas 25 (83%) exercised multiple univariate follow-up tests.

Clearly, instead of using MANOVA as a fully multivariate analysis, the majority of researchers in psychology and education use a MANOVA test even though their main interest is to study group mean differences on single outcome variables (Bird & Hadzi-Pavlovic, 2014). As Keselman et al. (1998) noted, "Focusing on results of multiple univariate analyses preceded by a MANOVA is no more logical than conducting an omnibus ANOVA but focusing on results of group contrast analyses" (p.16).

Procedures

In scrutinizing the statistical methods used in the articles, it was critical first to determine what specific methods to code. For guidance, previous research reviews in both the education and psychology fields were consulted (e.g., Keselman et al., 1998; Kieffer et al., 2001). In this systematic review, we focused on follow-up statistical procedures used in a MANOVA context. For the follow-up procedures, we recorded every follow-up test that the author(s) reported in the study's data analysis and results sections only. That is, any post hoc techniques that were mentioned in the literature review, discussion, or conclusion sections were not recorded. If the post hoc procedure was reported multiple times in a single study, it will be counted only once in our results.

Finally, if a multivariate test was used in accordance with a univariate one, it will be recorded as a multivariate follow-up and the accompanied use of univariate tests will be mentioned in the results. Since many studies use the term "univariate" for follow-up or post hoc tests, and such a term may include different tests, such as the Tukey, t-test, and Bonferroni correction, it will be regarded as a univariate method even though the specific test was not reported in the article. However, for the several authors that used more than one procedure within the same study, all of the reported procedures will be recorded under that same study.

In carrying out the review, we consulted the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2011), as reported in Bearman et al. (2012), and based the process on the nine phases of systematic reviews by David Gough (2007), director of the EPPI-Centre. We examined articles from five major databases that regularly publish research in education. These databases were selected according to the University of Connecticut's library website rating system and include ERIC, PsychINFO, Professional Development Collection, Academic Search Premier, and Teacher Reference Center (Education Research Databases, n.d.). We conducted the review through databases so we could examine a broader scope of topics utilizing MANOVA. The inclusion criteria were developed to review only empirical journal articles in education for both k-12 and post-secondary levels, reported in peer-reviewed journals, written in English, and published in the past five years (between January 2013 and March 2018).

In coming up with the search terms, we used more general keywords that appear anywhere in the article, such as "MANOVA," "multivariate analysis of variance," and/or "education," as the objective of this review is to have a varied collection of articles and to include any study that applied MANOVA. Although the focus of this systematic review is on studies in the education field, the final pool of articles included a variety of disciplines, such as medical, marketing, business, and sports. However, only those that reported follow-up tests or post hoc procedures are included in the literature review (n= 235). Figure 1 demonstrates a summary of the article review process.

The authors used Zotero software to collect, store, organize, and analyze the articles. This software allowed the researchers to manage the information, take notes, and generate references and bibliographies, citations, and reports (www.Zotero.org). The 344 articles were examined by the first author and then reviewed and checked by the second author of this study. The two coders discussed the results and solved any differences or disagreements that occurred. Both authors are PhD students, hold master's degrees in educational research and evaluation, and have mastered most of the common univariate and multivariate statistical methods in education.

Results

Figure 1 shows that out of 677 articles, we identified 424 articles that reported MANOVA somewhere in the study. A total of 344 empirical articles were determined to be eligible for review. Of these, 235 studies (68%), from 177 journals, actually reported follow-up procedures, whereas 109 studies (31%) did not report any follow-up statistical methods. Those that reported follow-up tests are analyzed further based on the five areas below.

Univariate post hoc

By far, the most common univariate post hoc tests reported were the multiple ANOVAs. Sixty articles (26%) reported its use as the only post hoc and another 51 studies (22%) reported its use with other post hoc procedures, such as the Tukey HSD and t-test. Thirty-five articles (15%) implied the use of a univariate F with either a post hoc test or a multiple comparison procedure (MCP). In total, approximately 62% of the
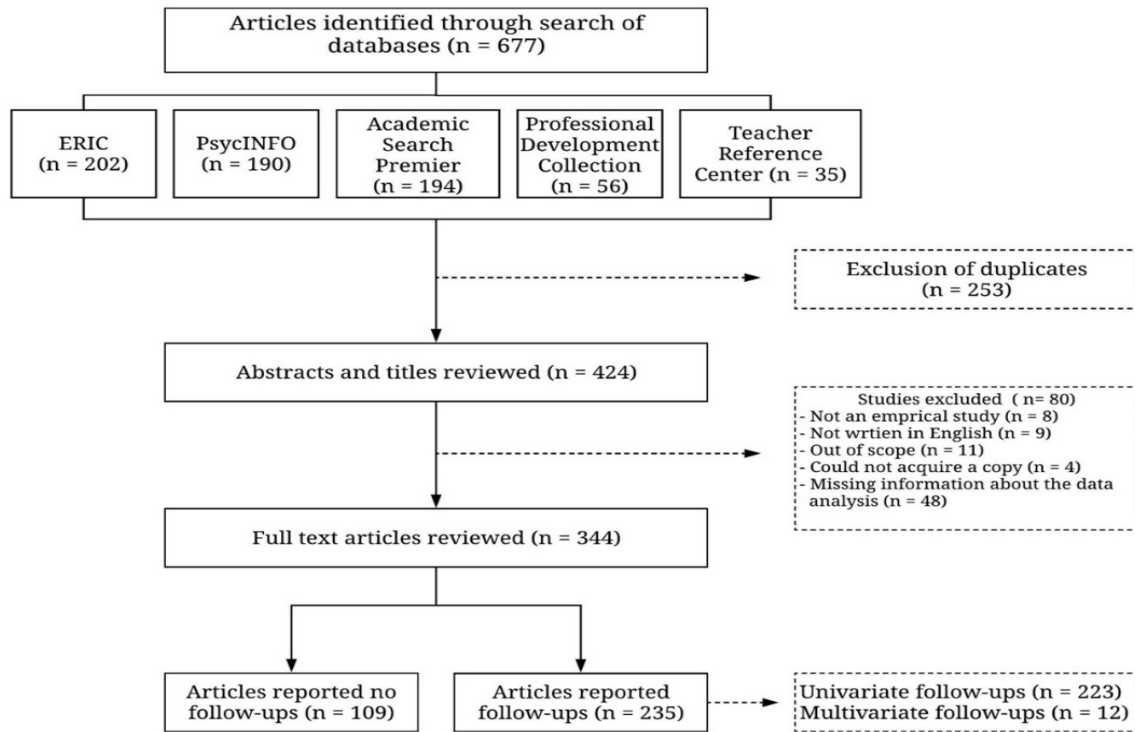
Figure 1. Summary of article review process.

Table 1. Frequency of univariate follow-up tests as reported.

Univariate Post hoc		Frequency	Percentage
ANOVAs		60	25%
ANOVA implied (univariate F test)		17	7%
Univariate F tests + post hoc + MCP		18	8%
	Bonferroni	24	10%
	Tukey HSD	13	6%
ANOVAs +	t-tests	6	3%
	Scheffé	4	2%
	Fisher's LSD	2	.9%
	Sidak	2	.9%
	Total	146	63%
Bonferroni		16	7%
Tukey HSD		15	6%
t-tests		8	3%
Scheffé		4	2%
Fisher's LSD		4	2%
Post hoc tests or MCP		14	6%
Games-Howell		4	2%
Duncan's T-tests		3	1%
Tamhane's T2		2	.9%
Dunnett C		2	.9%
Holm-Bonferroni + MPC		2	.9%
Helmert pre-planned comparisons		1	.4%
Fisher Z		1	.4%
Newman-Keuls		1	.4%
		77	33%
	Total	223	95%

studies (146 articles out of 235) that reported follow-up tests used ANOVA to explain MANOVA significant effects. Bonferroni procedures and the Tukey HSD were reported as the sole post hoc tests in 16 articles (7%) and 15 articles (6%), respectively. Meanwhile, t-tests and the Scheffé test were reported as the only post hoc in eight articles (3%) and four articles (2%), respectively. Fisher's LSD, Games-Howell, Tamhane's T2, and other post hoc tests were reported less frequently in a couple of articles, as is shown in Table 1. For the rest of the articles, the author(s) stated that they ran MCP and/or post hoc tests following significant results without specifying exactly what procedures were performed.

Multivariate post hoc procedures

The most common multivariate post hoc test in following-up MANOVA was Discriminant Function Analysis (DFA). This occurred in five articles as the only follow-up test and in five more articles accompanied by a series of ANOVAs, representing 5% of the total studies reporting post hoc tests. Hierarchical regression and Roy-Bargmann step-down tests were each reported once in two different studies (representing .8%). Table 2 shows the number and percentage of each multivariate post hoc procedure.

The 10 studies that reported using DFA indicated that the first discriminant function explained most of the underlying dimension. Different MANOVA designs were used: one- way (one study), factorial (four studies), three-way (one study) and four-groups MANOVA (four studies). The first discriminant function (1st DF) was the only DF reported in the one-way study whereas in three of the studies conducted factorial MANOVA the second discriminant function (2nd DF) was also reported. The last study that conducted factorial MANOVA reported all of the DFs. Similarly, the one study conducted three-way MANOVA reported the 1st DF only while the four studies utilized four-groups MANOVA reported 1st DF and 2nd DFs.

Table 2. Frequency of multivariate follow-up tests as reported.

Multivariate follow-up test	Frequency	Percentage
Discriminant Analysis	5	2%
Discriminant Analysis + ANOVAs	5	2%
Hierarchical Regression	1	.4%
Roy-Bargmann step-down test	1	.4%
Total	12	5%

Multivariate test statistics

In the majority of the articles, multivariate test statistics were reported for the significant results. As Table 3 shows, the most frequent multivariate tests used in reporting MANOVA results were Wilks' Lambda with 107 studies (46%), followed by Pillai's Trace with 44 studies (19%), and Hotelling's Trace with eight studies (3%). Fifty-eight studies (25%) reported the F test and five studies (2%) reported the p value, while seven other studies (3%) did not report any statistics for a MANOVA test. There were three studies that combined both Wilks' Lambda and Pillai's values and two other studies that provided tables of all four tests instead of reporting them within the results description. The remaining article indicated the use of the Huynh–Feldt test when reporting a MANOVA test result.

Table 3. Frequency of multivariate tests as reported.

Multivariate Tests	Frequency	Percentage
Wilks' Lambda	107	46%
Pillai's Trace	44	19%
Both (Wilks' + Pillai's)	3	1.5%
Hotelling's Trace	8	3%
Roy's Largest Root	0	0
All four in a Table	2	.9%
F tests Only	58	25%
No tests reported	7	3%
p value only	5	2%
Huynh–Feldt	1	.4%
Total	235	100%

Homogeneity tests

Levene's test is often used to determine the assumption of homogeneity of variance when conducting a t-test or ANOVA (Field, 2013), whereas Box's M test is used to determine the multivariate normality and whether variance-

Table 4. Frequency of homogeneity tests as reported.

Homogeneity tests	Frequency	Percentage
Box's M Tests	46	57%
Levene's Tests	35	43%
Total	81	100%

covariance matrices are equal for MANOVA (Huberty & Olinjeck, 2006). Table 4 shows the number and percentage of each homogeneity test reported. Out of 235 studies, only 81 reported the univariate and multivariate homogeneity test values. Box's M test was reported more frequently than Levene's test, with 46 studies (57%) and 35 studies (43%), respectively.

Software packages

The AERA (2006) and APA (2010) manuals of standards of statistical reporting urge researchers to report the statistical packages used, especially in multivariate analyses, such as MANOVA, regression analyses, and structural equation modeling analyses. As Table 5 shows, only 100 studies (43%) reported their use of statistical packages, utilizing such software as R, SAS, and Mplus. However, SPSS was the most frequently used software package (88 studies, or 88%) of those reported.

Table 5. Frequency of software packages as reported.

Homogeneity tests	Frequency	Percentage
SPSS	88	88%
SAS	1	1%
R	3	3%
AMOS	2	2%
PASW	3	3%
Mplus	1	1%
LISERL	1	1%
NetMiner	1	1%
Total	100	100%

Discussion

In this study, we examined the follow-up procedures used in a sample of published articles. As Tables 1 and 2 show, univariate follow-up methods are found overwhelmingly in the majority of articles. Generally, this trend coincides with what previous reviews found regarding statistical methods applied in both the education and psychology fields (e.g., Armstrong & Henson, 2005; Barton et al., 2016; Huberty & Morris, 1989; Keselman et al., 1998; Kieffer et al., 2001; Tonidandel & LeBreton, 2013; Warne et al., 2012). In concurrence with these reviews, education researchers seem to rely more on univariate post hoc procedures than multivariate ones. This general trend continues to occur regardless of the article's discipline. Though this trend is the most common, some researchers do apply multivariate follow-up procedures and seem to base their decisions on multivariate principles.

Other Findings

While conducting this review, some inclinations among the researchers were observed. The most important observation concerned how researchers reported the resultant effect sizes. The AERA (2006) and APA (2010) manuals of standards of statistical reporting indicate that researchers should always report effect size indices. Surprisingly, a great portion of the MANOVA articles reported the η2, yet they simply based the magnitude of their effect sizes on Cohen's (1988) guideline (small, medium, or large) without further explanation. In practice, a large magnitude can indicate an insignificant effect, whereas a small magnitude can be a major indication of a significant effect (Durlak, 2009), though few authors articulated their understanding of the effect sizes reported (Thompson, 2006). In some cases, the authors explained their use of univariate methods as a result of an insufficient sample size required for the multivariate post hoc. Others indicated the use of either method is contingent on the purpose of the post hoc test (i.e. explorative versus confirmatory purposes). Also, some explained that DFA is used when there are more than two groups. Although such practices are known to be appropriate, it does not negate the fact that a multivariate technique is the legitimate method for clarifying multivariate effects.

Limitations

Several limitations are evident in this review. First, the MANOVA procedures used in the analyzed studies were not examined in terms of their appropriateness to the posed research questions or the data collected (e.g., sample size). That is, the review was mainly contextual and no judgment was made about fulfilling the assumptions required for the analysis. Second, the system we established to code the articles' information did not always correspond seamlessly with what was reported. In many cases, for example, where the follow-ups were just implied, we often make a subjective judgment about whether the authors conducted univariate or multivariate follow-ups. Having two coders for examining the articles, however, made these judgments reasonable. Furthermore, it may be that we missed other research that applied multivariate methods and was reported in other journals not included in our sample. Finally, a time period

of five years may be too short to observe more multivariate post hoc practices that are recommended by the theoretical discussion presented in the literature.

Conclusion

Overall, the present review of the literature reveals that researchers in both education and psychology share with researchers in other disciplines, such as medicine, sport, and business, a common practice in their use of univariate methods. This review can provide guidance on what areas novel researchers and practitioners need more training or instructions on. Also, it would assist them in choosing from the appropriate statistical methods applied in the field. Generally speaking, researchers support conventional univariate post hoc tests over multivariate methods for analyzing the significant group mean difference observed in their multivariate data. However, a small portion of articles used multivariate follow-up methods (primarily DFA). We believe that researchers in the field should be knowledgeable of the complex multivariate methods to embrace the complexity of social phenomena. It is hoped that the findings of this review will encourage a methodological confab among education researchers, which would assist researchers in carrying out more vigorous research and in gaining better insights into social problems.

References

Education Research Databases. (n.d.). Retrieved February 6, 2018, from
http://rdl.lib.uconn.edu/subjects/1889

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*, 33-40. doi:10.3102/0013189x035006033

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

Armstrong, S. A., & Henson, R. K. (2005). Statistical practices of IJPT researchers: A review from 1993-2000. *International Journal of Play Therapy*, *14*(1), 7.

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 discriminate analysis. *Research quarterly for exercise and sport*, *87*(4), 365-375.

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, *31*(5), 625-640.

Bird, K. D., & Hadzi-Pavlovic, D. (2014). Controlling the maximum familywise Type I error rate in analyses of multivariate experiments. *Psychological Methods*, *19*(2), 265.

Borgen, F. H., & Seling, M. J. (1978). Uses of discriminant analysis following MANOVA: Multivariate statistics for multivariate purposes. *Journal of Applied Psychology*, *63*(6), 689.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology*, *34*(9), 917-928.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research papers in education*, *22*(2), 213-228.

Higgins, J. P. (2011). Green S. Cochrane handbook for systematic reviews of interventions version 5.1. 0. *The cochrane collaboration*, *5*(0).

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological bulletin*, *105*(2), 302.

Huberty, C. J., & Olejnik, S. (2006). Applied MANOVA and discriminant analysis. *Wiley Series in Probability and Statistics*. doi:10.1002/047178947x

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*(3), 350-386.

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in AERJ and JCP articles from 1988 to 1997: A methodological review. *The Journal of Experimental Education*, *69*(3), 280-309.

Thompson, B. (1999). Common methodology mistakes in educational research, revisited, along with a Primer on both effect sizes and the Bootstrap.

Thompson, B. (2006). *Foundations of behavioral statistics*. New York, NY: Guilford Press.

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*(3), 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*.

Warne, R. T., Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly*, *56*(3), 134-149. doi:10.1177/0016986212444122.

Send correspondence to: Gordon Brooks
Ohio University
Email: brooksg@ohio.edu