MANOVA: A Procedure Whose Time Has Passed?

Francis L. Huang, PhD
University of Missouri

Abstract

Multivariate analysis of variance (MANOVA) is a statistical procedure commonly used in fields such as education and psychology. However, MANOVA's popularity may actually be for the wrong reasons. The large majority of published research using MANOVA focus on univariate research questions rather than the multivariate questions that MANOVA is said to specifically address. Given the more complicated and limited nature of interpreting MANOVA effects (which researchers may not actually be interested in given the actual post-hoc strategies employed) and that various flexible and well-known statistical alternatives are available, I suggest that researchers consult these better known, robust, and flexible procedures instead, given the proper match with the research question of interest. Just because a researcher has multiple dependent variables of interest does not mean that a MANOVA should be used at all.

Keywords: MANOVA, ANOVA, Type I error, SEM, regression

The multivariate analysis of variance (MANOVA) procedure is often used by education researchers (Keselman et al., 1998; Warne, 2014) and a wide range of social, cognitive, developmental, organizational, and clinical psychologists (Enders, 2003; Tonidandel & LeBreton, 2013). Among gifted education journals, from 2006 to 2010, almost a quarter of quantitative articles used MANOVA as well (Warne, Lazo, Ramos, & Ritter, 2012). However, surprisingly, despite its popularity, much of the analyses performed focus on answering univariate rather than true multivariate[1] questions (Keselman et al., 1998), which suggests that MANOVA should not have been used in the first place. In other words, MANOVA may be popular but for the wrong reasons. Although from a teaching perspective, introducing MANOVA as an extension of the more basic univariate ANOVA provides some instructional scaffolding building on student's prior knowledge, the questions that multivariate vs. univariate procedures answer are quite different (Enders, 2003; Huberty & Morris, 1989; Zientek & Thompson, 2009).

Considering that performing a MANOVA is prone to error (see Smith, Lamb, & Henson, 2020) and may not actually be the best method in answering particular types of research questions, I suggest that researchers review other suitable and accessible statistical procedures. Everitt and Hothorn (2011), in explaining why they had excluded MANOVA from their multivariate textbook, stated "we are not convinced that MANOVA is now of much more than historical interest" (p. vii). Other methodologists have relegated MANOVA to the "multivariate dustbin" (Ender, 2017) along with other less often used multivariate techniques (e.g., canonical correlation). I am not suggesting that MANOVA be shelved because it is old (e.g., look at factor analysis) but that researchers should consider alternative procedures. Using alternative statistical

[1] I refer to multivariate techniques as those that focus simultaneously on multiple dependent variables. For example, a regression with one dependent variable and multiple predictors can be referred to as a multiple regression or a multivariable technique but not a multivariate regression.

procedures is not uncommon and examples of this include using logistic regression instead of discriminant function analysis (Fan & Wang, 1999) or using regression with cluster robust standard errors instead of multilevel modeling (MLM; Huang, 2016). I argue that in many instances MANOVA should not be used because: 1) the actual research questions may not be multivariate to begin with; 2) other statistical methods can be better suited to answer a research question of interest; and 3) given the challenges and assumptions involved in appropriately performing and interpreting MANOVA results, researchers are better off using other less error-prone procedures. I discuss some of the issues with using MANOVA but also provide several alternatives.

What is the outcome variable in a MANOVA really?

Unlike traditional univariate techniques, multivariate methods such as MANOVA work by forming synthetic or artificial variables based on a linear combination of the measured variables. In the case of MANOVA, the measured DVs are weighted and added together to create new DVs, also referred to as variates or multivariable composites. The weights are derived, as part of the MANOVA procedure, in order to create a new DV that maximizes group differences and the new constructs/variates are formed not necessarily based on any particular theory. The test of statistical significance then focuses on whether the centroids of these variates differ among the groups being evaluated over and above what can be expected by chance.

For example, in a MANOVA example given by Grice and Iwasaki (2007), a researcher may be interested in how children in public and private schools differ on three tests that measure 1) reading, 2) mathematics, and 3) moral reasoning. When a MANOVA (or a Hotelling's T^2 with two groups) is performed, the outcome is a weighted linear combination of the three DVs—one measured DV may get more weight and contribute more to the variate while another measured DV may get much less weight, hardly contributing anything at all. The linear combination is what the MANOVA procedure determines and what should be interpreted (Warne, 2014). The multivariate research question asks about the difference between groups based on the variate (Zientek & Thompson, 2009), not specifically the measured variables. As indicated by Huberty and Olejnik (2006):

> A two-group MANOVA may be viewed as a (univariate) two-group ANOVA where the single outcome variable consists of a linear combination of the original multiple outcome variables. It is these variable combinations that are the center of attention…That is, the main reason for conducting a MANOVA/DDA is to interpret … the resulting variable combination(s) that is(are) associated with group differences (p. 7).

This is also a reason why researchers, on occasion, may not find statistically significant pairwise differences in the individual DVs even after finding a statistically significant multivariate effect. In essence, what is being evaluated is whether the groups differ on the variate, not the individual DVs. The multivariate research question focuses on whether groups differ on the variate (or the linear combination of the DVs). For Grice and Iwasaki (2007), their multivariate research question may be: "Do children in public and private schools differ on a variate formed using three measured constructs?"

Using MANOVA, if followed up appropriately[2] with a procedure like DDA, as recommended several authors (Smith et al., 2020; Warne, 2014), researchers describe the differences in the variates by group, based on the discriminant functions, investigating both the structure and standardized coefficients. Although the MANOVA research question focuses on whether groups differ on the variate, the DDA question focuses on how the groups differ on the variate. However, after conducting a DDA, researchers are not in a position to evaluate specific levels of mean differences in the DVs among the groups (e.g., Group A had higher Y_1 vs. Group B). As an example, Puryear and Kettler (2017), who conducted a one-way MANOVA investigating rural gifted education and proximity, stated in their limitations: "It is difficult to provide clear, meaningful "If X does this, Y does this" sorts of interpretations for specific variables. Thus, we have had to limit our conclusions to discussions of relative variable importance…" (p. 151). However, this is a limitation of the procedure used, not necessarily of the data. If the relative importance of variables is of interest or if the research question focuses on differences in group centroids, then a MANOVA and DDA should be fine.[3] If the interest is in determining and specifically quantifying the differences between groups using the DVs of interest (e.g., reading, mathematics), then performing a MANOVA is not really necessary.

Even with the selection of variables for the DVs, MANOVA is possible if the DVs exhibit a moderate to strong correlation (e.g., $r = \pm .60$) with each other, which allows for the formation of the linear composites (Tabachnick & Fidell, 2019). If the DVs are weakly or not correlated to each other, then it will not be possible to create meaningful linear composites and researchers should just conduct independent univariate tests. If, however, the variables show a strong relationship with each other, then testing latent mean differences using structural equation modeling (SEM) may be preferred (Cole, Maxwell, Arvey, & Salas, 1993). With SEM (Adelson, 2012), researchers have several additional benefits such as having an outcome devoid of measurement error, model fit indices, the ability to properly model categorical and multilevel data, and a straightforward interpretation of mean differences using the structural component of the model.

What has MANOVA been used for primarily?

In practice, after detecting a statistically significant multivariate effect, researchers frequently resort to univariate ANOVAs and multiple pairwise *t*-tests to explain what is accounting for these group differences. Three decades ago, Huberty and Morris (1989) indicated that 96% of MANOVA follow ups were univariate ANOVAs. In educational research, 84% of

---

[2] Another procedure, if there is a theoretical ordering of the DVs, is the Roy-Bargmann stepdown procedure (Finch, 2007). In a review of 64 articles that used MANOVA in gifted education research, only one article had used the Roy-Bargmann procedure (Warne et al., 2012). Others, though, do not recommend its use as a posthoc procedure (Finch, 2007). The Pituch and Stevens' (2015) multivariate textbook used to include a section on this procedure as well, but that section has been relegated to the online appendix.

[3] Toninandel and LeBreton (2013) indicated that interpreting the resulting standardized regression coefficients (alone) from a DDA is problematic because often, predictors are not orthogonal to each other. Instead, they recommend yet another procedure referred to as relative weight analysis.

studies followed that strategy as well (Keselman et al., 1998). A more recent review indicated that these numbers have hardly changed over time (Tonidandel & LeBreton, 2013). In gifted research, the most common follow up after a statistically significant MANOVA was a series of ANOVAs (Warne et al., 2012). The popularity of the multivariate-univariate approach is not surprising as several leading textbooks on multivariate methods have suggested this strategy (Pituch & Stevens, 2015; Tabachnick & Fidell, 2019).[4] In these cases, the MANOVA test functions as a gatekeeper and is used with the hopes of controlling for Type I error rates when analyzing multiple dependent variables (DVs).

Although the use of multiple pairwise comparisons is a natural follow up given researchers are more likely familiar with ANOVAs and some software provide these ANOVAs by default when conducting a MANOVA, several authors have specifically warned against this practice (Grice & Iwasaki, 2007; Keselman et al., 1998; Smith et al., 2020; Warne, 2014). Several methodologists have clearly opposed this two-step multivariate-univariate process (Enders, 2003). Huberty and Morris (1989) wrote:

> Even though it is a fairly popular analysis route to take in the behavioral sciences, conducting a MANOVA as a preliminary step to multiple ANOVAS is not only unnecessary but irrelevant as well. We consider to be a myth the idea that one is controlling Type I error probability by following a significant MANOVA test with multiple ANOVA tests, each conducted using conventional significance levels. Furthermore, the research questions addressed by a MANOVA and by multiple ANOVAS are different; the results of one analysis may have little or no direct substantive bearing on the results of the other. To require MANOVA as a prerequisite of multiple ANOVAS is illogical, and the comfort of statistical protection is an illusion. (p. 307)

Other authors have suggested that an appropriate follow up procedure after finding a statistically significant MANOVA test is to conduct a descriptive discriminant analysis (DDA) (see Enders, 2003; Smith et al., 2020; Warne, 2014). A DDA research question focuses on what variables contribute to the difference in the variates between the groups. The standardized discriminant and structure coefficients are then interpreted as part of a DDA.

However, if a researcher is simply interested in answering a research question such as "How much higher did the treatment group participants score on outcomes Y_1, Y_2, and Y_3, compared to the control group participants", a DDA does not provide that type of information. If, however, researchers are interested in quantifying differences between groups (which is typical in randomized control trials or experiments), then researchers are better off conducting basic regressions/*t*-tests and then performing a Bonferroni adjustment or, even more powerful yet, a Benjamini-Hochberg (1995) correction instead (see Institute of Education Sciences, 2014, pp. G.2 - G.5 which provides a basic tutorial on this). Additionally, the added benefits of using a regression with dummy-coded grouping variables (over a basic t-test) are the ability to include relevant covariates without adding much complexity to the model and that the results are easy to interpret.

---

[4] Although covering much more than MANOVA, according to Google Scholar, Tabachnick and Fidell's (2019) multivariate textbook has been cited more than 90,000 times as of August 2019.

What about MANOVA for Repeated Measures?

MANOVA can also been used to analyze repeated measurements taken from individuals over time (Hahs-Vaughn, 2016; Hair, Black, Babin, & Anderson, 2009). However, the analyses of repeated measures has been largely superseded by growth models or linear mixed effects models (i.e., MLMs; Everitt & Hothorn, 2011; Hahs-Vaughn, 2016). The use of MLM for the study of growth over time is well documented (Singer, 1998), handles missing data well (Luke, 2004), works better than MANOVA when the data are unbalanced (Schuster & Lubbe, 2015), and is a highly flexible procedure (Singer & Willett, 2003).

Other Considerations with MANOVA

Although researchers typically have to make decisions when analyzing data regardless of statistical method used (e.g., What rotation to use in a factor analysis? What method to use to adjust for multiple comparisons?), when performing a MANOVA, there are various decisions at each stage of the analysis with no clear best choice. For example, multiple statistics (e.g., Wilk's Λ, Roy's θ) can be used for the multivariate test where there is no superior choice which can confuse researchers (Haase & Ellis, 1987).[5] Others indicate that the choice of test used should be guided by the conditions being studied (Hahs-Vaughn, 2016). In addition, there are a "bewildering array of follow-up tests that are designed to assess the relative importance of the individual dependent variables" (Haase & Ellis, 1987, p. 410). Popular multivariate textbooks often suggest the simpler univariate follow-ups (Pituch & Stevens, 2015; Tabachnick & Fidell, 2019), but then several articles oppose this practice (Enders, 2003; Huberty & Morris, 1989; Smith et al., 2020; Warne, 2014). In many instances though, when faced with the necessary choice, researchers are likely to default to the simpler multivariate-univariate approach because that is commonly done by the majority of researchers using MANOVA.

As part of the general linear model (GLM), MANOVA still has to satisfy certain model assumptions such as observation independence. In a MANOVA context, the violation of observation independence is very serious (Pituch & Stevens, 2015). For regressions for example, using multilevel models is a well-known solution to the issue of observation independence resulting from nesting (Huang, 2016). With MANOVA though, not many alternatives for violating this assumption have been suggested. Some have suggested using the groups as the units of analysis and analyzing group means instead of individual responses or by using a decreased alpha (e.g., .01 vs. .05) level (Hahs-Vaughn, 2016; Pituch & Stevens, 2015). An alternative procedure though, to completely remove group effects resulting from clustering, is to demean (or group center) the data (Huang, 2016) and then perform the MANOVA. Pituch and Stevens (2015) also discuss multivariate multilevel modelling (MVMM) although the outcomes of interest are the individual DVs, not a weighted linear combination of the DVs.

Another well-known MANOVA model assumption is the homogeneity of covariance matrices (though with relatively similar group sizes, MANOVA is robust to this violation) (Tabachnick & Fidell, 2019). Box's M test of the equality of covariance matrices is often used to

[5] In instances when there are only two groups, results will be the same regardless of statistic used.

test this assumption (Friendly & Sigal, 2018). However, given the number of dependent variables and groups, this assumption may not be met, especially with large data sets and unequal group sizes. Other researchers suggest inspecting the log determinants to see whether they are in the 'same ballpark' (Huberty & Olejnik, 2006) which is highly subjective. Still others provide guidance that as long as the variance ratio between the largest and smallest groups is not beyond 10:1, then this violation is not considered problematic (Tabachnick & Fidell, 2019). However, it is not clear what the basis of these rules of thumb are and no references are provided to support these assertions. A possible area for future study may be the use of Monte Carlo simulations investigating the tenability of the different rules.

Conclusions

I provide some practical recommendations to applied researchers who are thinking of using MANOVA (see Table 1). If a researcher is interested in quantifying differences between groups (e.g., how do the *Y*s change based on group assignment), then multiple *t*-tests, ANOVAs, or regressions can be used together with a Bonferroni correction or a Benjamini-Hochberg procedure instead to control for Type I error. If a research question involves latent mean differences, SEM can be used. If analyzing growth over time, an MLM or growth curve model provides yet another well-established alternative. Although MANOVA as a basic multivariate procedure may have some historical interest, if the research questions do not really focus on the variate, then analysts are better off using more well understood statistical procedures. The analytic technique used should match the research question. Given that the large majority of questions may actually be univariate (rather than multivariate) in nature, the popularity of MANOVA as a procedure of choice is not warranted, especially when there are a number of acceptable alternatives. As stated by Hancock (2019), "I don't know what the question is, but the answer is almost never 'MANOVA.'"

Table 1.
Alternatives to MANOVA.

Research Focus	Procedure	Advantage
Mean differences in dependent variables by group membership	Regression with correction for multiple comparisons	• Commonly used and well understood • Corrects for Type I errors
Latent mean differences by group membership	Structural equation modeling	• Makes use of intercorrelations among variables • Outcome is devoid of measurement error • Model fit indices are provided
Change over time	Growth curve modeling or multilevel modeling	• Can handle missing data or uneven occurrence of events

Notes. All alternative procedures can also include covariates and can account for clustered data.

References

Adelson, J. L. (2012). Examining relationships and effects in gifted education research: An introduction to structural equation modeling. *Gifted Child Quarterly*, *56*, 47–55.

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. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, *114*, 174–184. https://doi.org/10.1037/0033-2909.114.1.174

Ender, P. (2017). The multivariate dustbin. In *2017 Stata Conference*. Stata Users Group. Retrieved from https://www.stata.com/meeting/baltimore17/slides/Baltimore17_Ender.pdf

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

Everitt, B., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. New York, NY: Springer Science & Business Media.

Fan, X., & Wang, L. (1999). Comparing linear discriminant function with logistic regression for the two-group classification problem. *The Journal of Experimental Education*, *67*, 265–286. https://doi.org/10.1080/00220979909598356

Finch, W. H. (2007). Performance of the Roy-Bargmann stepdown procedure as a follow up to a significant MANOVA. *Multiple Linear Regression Viewpoints*, *33*, 12–22.

Friendly, M., & Sigal, M. (2018). Visualizing tests for equality of covariance matrices. *The American Statistician*, *0*, 1–12. https://doi.org/10.1080/00031305.2018.1497537

Grice, J. W., & Iwasaki, M. (2007). A truly multivariate approach to MANOVA. *Applied Multivariate Research*, *12*, 199–226. https://doi.org/10.22329/amr.v12i3.660

Haase, R. F., & Ellis, M. V. (1987). Multivariate analysis of variance. *Journal of Counseling Psychology*, *34*, 404–413. https://doi.org/10.1037/0022-0167.34.4.404

Hahs-Vaughn, D. L. (2016). *Applied multivariate statistical concepts*. New York, NY: Routledge.

Hair, J., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Hancock, G. [GregoryRHancock]. (2019, October 1). I don't know what the question is, but the answer is almost never "MANOVA." [Tweet] Retrieved from https://twitter.com/GregoryRHancock/status/1179006872688496640

Huang, F. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education*, *84*, 175–196. https://doi.org/10.1080/00220973.2014.952397

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

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (Vol. 498). Hoboken, NJ: John Wiley & Sons.

Institute of Education Sciences. (2014). What Works Clearinghouse: WWC procedures and standards handbook version 3.0. Retrieved September 10, 2015, from https://ies.ed.gov/ncee/wwc/docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., … Keselman, J. C. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68,* 350–386. https://doi.org/10.3102/00346543068003350

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.

Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*. New York, NY: Routledge.

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

Schuster, C., & Lubbe, D. (2015). MANOVA versus mixed models: Comparing approaches to modeling within-subject dependence. In M. Stemmler, A. von Eye, & W. Wiedermann (Eds.), *Dependent Data in Social Sciences Research* (Vol. 145, pp. 369–385). Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-20585-4_16

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *23*, 323–355. https://doi.org/10.2307/1165280

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford university press.

Smith, K., Lamb, K., & Henson, R. (2020). Making meaning out of MANOVA: Using descriptive discriminant analysis for multivariate post hoc testing in gifted education research. *Gifted Child Quarterly, 64. xxx*

Tabachnick, B. G., & Fidell, L. (2019). *Using multivariate statistics* (7th edition). New York, NY: Pearson.

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. https://doi.org/10.1037/a0032001

Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists. *Practical Assessment, Research & Evaluation*, *19*. Retrieved from https://pareonline.net/getvn.asp?v=19&n=17

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. https://doi.org/10.1177/0016986212444122

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