and quasi-experimental designs for research. Chicago: Rand McNally.

Cook, T.D. & Reichardt, C.S. (1979). *Qualitative and quantitative methods in evaluation research.* Beverly Hills, CA:Sage.

Creswell, J.W. (1994) *Research design: Qualitative and quantitative approaches.* Thousands Oaks, CA:Sage.

Denzin, N.K. (1994). Evaluating qualitative research in the poststructural moment:The lessons James Joyce teaches us. *Qualitative Studies in Education,7*(4), 295-308.

Eisner, E.W & Peshkin, A. (Eds.) (1990). *Qualitative inquiry in education: The continuing debate.* New York: Teachers College Press.

Gay, L.R. (1987). *Educational research:Competencies for analysis and application.* (3rd ed.) Columbus, OH:Merrill.

Gueulette, C.M., Newgent, R.A., & Newman, I. (2000, Feb.) *Qualitative research:The illegitimatization of intent and purpose.* Paper presented at the Association for Advancement of Educational Research annual meeting, Ponte Verde, FL.

Hoshmond, L. (1999). Locating the qualitative research genre. In M. Kopala & L.A. Suzki (Eds.) *Using qualitative methods in psychology.* (pp. 15). Thousand Oaks, CA: Sage.

LeCompte, M.D., & Goetz, J.P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research, 52*(1), 31-60.

Lincoln, Y.S., (1990). Toward a categorical imperative for qualitative research. In E.W. Eisner & A. Peshkin (Eds.) *Qualitative inquiry in education: The continuing debate* (pp.277-295). New York: Teachers College Press.

Lincoln, Y. S., (1995). Emerging criteria for quality in qualitative research. *Qualitative Inquiry, 1* (3), 275-289.

Lincoln, Y. S. & Guba, E.G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Miles, M.B., & Huberman, A.M. (1984). Drawing valid meaning from qualitative data:Toward a shared craft. *Educational Researcher, 13*(5), 20-30.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook (2nd ed.)* Thousand oaks, CA: Sage.

Newman, I. & Benz, C. (1998) *Qualitative-quantitative research methodology: Exploring the interactive continuum.* Carbondale, IL: Southern Illinois University Press.

Newman, I. & Newman, C. (1994). *Conceptual Statistics for Beginners.* Washington, D.C:University Press of America.

Patton, M.Q. (1990) *Qualitative evaluation and research methods (2nd ed.).* Newbury Park, CA: Sage.

Reichardt, C.S., & Cook, T.D. (1979). Beyond qualitative versus quantitative methods. In T.D. Cook & C.S. Reichardt (Eds.) *Qualitative and quantitative methods in evaluation research* (pp. 7-32). Beverly Hills, CA:Sage.

Schratz, M. (Ed.). (1993). *Qualitative voices in educational research.* London: Palmer Press.

Spradley, J.P. (1979). *The ethnographic interview.* New York: Holt Rinehart& Winston.

Tashakkori, A. & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousands Oaks, CA: Sage.

Isadore Newman is a Distinguished Professor at the University of Akron.

Carole Newman is an Associate Professor at the University of Akron.

A Three-Step Adjustment Procedure for Type I Error Rates

Isadore Newman
The University of Akron

John Fraas
Ashland University

John M. Laux
The University of Akron

Research and program evaluators routinely implement alpha adjustment procedures for multiple statistical tests of the differences among group means. Ryan (1959) and Stevens (1996) suggest, however, that multiple statistical tests are encountered in many studies in which the researchers and program evaluators do not readily identify the need to adjust for inflated Type I error rates. These types of studies include the following: (a) ANOVA designs that contain main effects and interaction effects, (b) analyses of multiple dependent variables, (c) analyses of multiple regression models, and (d) analyses of numerous correlation coefficients.

Even when researchers are aware of the need to adjust for inflated Type I error rates, they must choose from a variety of alpha adjustment procedures with each procedure having its own assumptions and degree of complexity (Hays, 1988; Kirk, 1982, 1994; Stevens, 1996; Toothaker, 1991; Winer, Brown, & Michels, 1991). As noted by Kirk (1994):

> The selection of an appropriate multiple comparisons procedure [an alpha adjustment procedure] is a complex task.... Unfortunately, as research tools become more and more complex, fewer and fewer researchers understand the assumptions and limitations of the tools. We can hope that the next methodological breakthroughs will result in a return to conceptual simplicity and away from increasing complexity. (pp. 117-118)

In an attempt to provide conceptual simplicity and a movement away from complexity, we propose that researchers and program evaluators contemplate utilizing a three-step procedure when adjusting the alpha levels in any type of study that contains multiple statistical tests, including the types of studies mentioned by Ryan (1959) and Stevens (1996).

In addition to being a simple and a flexible method for adjusting Type I error rates, we believe our three-step adjustment procedure appropriately requires researchers to reflect on the rationale used to make those adjustments. Specifically, this procedure requires the researchers to address three fundamental questions when adjusting the alpha levels in studies that contain multiple statistical tests:

1. What conceptual error rate unit or units are contained in the study?
2. What are the number and nature of the statistical tests contained in a given error rate unit?
3. What specific adjustment should be made to the alpha level of each statistical test contained in a given error rate unit in order to maintain the desired overall alpha level for that unit?

The use of this three-step procedure should provide researchers, and hopefully the readers of their findings, a more complete understanding of the rationale used to statistically test the data.

Adjusting Alpha Levels for Multiple Statistical Tests

The use of multiple statistical tests may lead to a situation where the chance of committing at least one Type I error, i.e., a situation in which a true null hypothesis is rejected, is significantly increased. To understand the impact of multiple statistical tests on the probability of committing at least one Type I error, assume that the researchers identified each statistical test as its own unit of error. If a statistical test is conducted with a selected alpha level of .05, the probability of committing a Type I error is .05 for that test. For a unit that contains two statistical tests, the probability of at least one Type I error approaches .10 when the alpha level is set at .05 for each test. The probability of committing at least one Type I error for a specified number of orthogonal statistical tests with the alpha value for an individual test set at a specified alpha level can be calculated as follows:

$$p(\text{at least one Type I error}) = 1 - (1 - \alpha_{ind})^m \; m\alpha_{ind} \qquad (1.1)$$

where:

1. The symbol m represents the number of specific statistical tests being conducted.
2. The symbol α_{ind} represents the alpha level established for each statistical test.
3. The approximate upper bound on the probability level of committing at least one Type I error is represented by $m\alpha_{ind}$, i.e., m multiplied by $_{ind}$.

If the m statistical tests are not orthogonal, the probability of committing at least one Type I error can be summarized as follows:

$$p(\text{at least one Type I error}) \leq 1 - (1 - \alpha_{ind})^m \leq m\alpha_{ind} \qquad (1.2)$$

Empirical research indicates the probability of committing at least one Type I error is fairly close to the value indicated by $1 - (1 - \alpha_{ind})^m$ for statistical tests that are not orthogonal (Toothaker, 1991).

As revealed by Equations 1.1 and 1.2 and as noted by Hays (1988), Kirk (1982), Stevens (1996), Toothaker (1991), and Winer, Brown, and Michels (1991), the chance of committing a Type I error can increase dramatically when multiple statistical tests are conducted. The previously-mentioned authors provide detailed presentations of numerous methods that can be used to adjust the Type I error rates for multiple statistical tests. The proposed adjustment procedures include the following: (a) Fisher's least significance differences test (Fisher, 1949, pp. 56-58); (b) Tukey's HSD test (Tukey, 1953); (c) Spjøtvoll and Stoline's modification of the HSD test (Spjøtvoll & Stoline, (1973); (d) Tukey-Kramer modification of the HSD test (Tukey, 1953; Kramer, 1956); (e) Scheffé's test (Scheffé, 1953); (f) Brown-Forsythe BF procedure (Brown & Forsythe, 1974); (g) Newman-Keuls test (Newman, 1939; Keuls, 1952); (h) Duncan's new multiple range test (Duncan, 1955); (i) Bonferroni-type adjustment procedure (Dunn, 1961; Newman & Fry, 1972); and, (j) modified Bonferroni-type adjustment procedures (Holm, 1979; Holland & Copenhaver, 1987; Hockberg, 1988).

As indicated by the number of possible multiple test adjustment procedures which have been developed, the selection and implementation of an adjustment technique may appear to be a daunting task for a researcher or program evaluator. In an attempt to heed Kirk's (1994) call for a return to conceptual simplicity and a decrease in complexity, we are proposing that researchers and program evaluators consider a simple three-step procedure when adjusting alpha levels in studies containing multiple statistical tests. To demonstrate the use of our proposed three-step alpha adjustment procedure, we have applied it to various research scenarios in a hypothetical study.

Hypothetical Study

The hypothetical study used to illustrate the application of our proposed three-step adjustment procedure involves the analysis of teacher efficacy levels. In this study the researchers attempted to gauge the differences in teacher efficacy levels of teachers exposed to three types of in-service seminars. The following two teacher efficacy constructs were measured in the study: (a) teaching efficacy, and (b) personal efficacy. These two teacher efficacy constructs, which served as the two dependent variables in the study, were analyzed through two traditional three-by-two analysis of variance (3X2 ANOVA) procedures, with the type of seminar completed by the educator (treatment) and gender serving as the independent variables. The use of two separate two-way

ANOVAs requires the researchers to conduct as many as six statistical tests on each of the two dependent variables. These statistical tests consist of the following for each dependent variable:

1. Two statistical tests of the two main effects (treatment and gender).
2. One statistical test of the interaction effect.
3. Three follow-up statistical tests of the three pairwise comparisons of the three treatment means.

To illustrate how researchers could make different adjustments to the alpha levels for the tests contained in this hypothetical study, assume three groups of researchers approached this study from three different perspectives. The three different perspectives, which are referred to as scenarios, are as follows:

1. Scenario 1 -- In this scenario the researchers consider the two dependent variables to be part of one overall analysis. In addition, the researchers have no preconceived notion of what the various statistical test results will reveal. Thus, the researchers are approaching the study from an exploratory perspective, i.e., they lack a strong theoretical base or previous research results from which to predict the results of the statistical tests being conducted. We will refer to these types of statistical tests of hypotheses, which are not based on theory or previous research, as exploratory in nature.

2. Scenario 2 -- In this scenario we assume the researchers consider the two dependent variables as constituting two separate analyses in the same study or two separate studies. In addition, we assume, as is the case under Scenario 1, the researchers have no preconceived notion of what the various statistical test results will reveal. That is, the study contains exploratory statistical tests.

3. Scenario 3 -- Similar to Scenario 2, we assume the researchers in Scenario 3 considered the analyses of the two dependent variables as constituting two separate analyses. Unlike the researchers in Scenarios 1 and 2, however, we assume the researchers predicted that interaction effects would exist between the treatment and the gender variables for both of the dependent variables. Since the researchers were willing to predict the existence of interaction effects based on theory and/or the results of previous research, the statistical tests of the interaction effects are not exploratory in nature. Rather, we will label them as theory based. If the interaction effects are found not to be statistically significant and the main effects are tested, the statistical tests of those main effects, however, will be considered as exploratory.

The Three-Step Adjustment Procedure

We take the position that controlling the Type I error rates in studies that contain multiple statistical tests, such as the hypothetical study previously presented, can be accomplished in a conceptually simple manner by implementing a three-step adjustment procedure. First, the researchers need to identify the rate error units contained in the study. Second, the researchers must determine the number and nature of statistical tests contained in each error rate unit. Third, the researchers must implement a procedure that will adjust the alpha levels of the statistical tests contained in the error rate unit. To illustrate how this three-step procedure can be utilized by researchers, we have applied it to each of the three research scenarios previously discussed.

Step One: Defining the Error Rate Units

In order to adjust the Type I error rates for a study that involves multiple statistical tests, the researchers must first specify the error rate unit. As noted by a number of authors, the relative merits of using one conceptual error unit over another can be debated (Duncan, 1955; McHugh & Ellis, 1955; Ryan, 1959, 1962; Wilson, 1962). Nevertheless, the identification of the error rate units, which we find is seldom discussed or even presented in research studies that include multiple statistical tests, is an important element in understanding the context in which a study's statistical tests were conducted.

Since a given error rate unit forms the logical framework on which the adjustments of the alpha levels of the individual statistical tests are based, researchers should be able to identify each error rate unit they believe exists in the study. It would also be helpful to the readers of such research if the researchers would state the logic on which the unit was based. How these error units are defined is heavily dependent on the researchers' conceptualization of their study. For example, a researcher who is engaged only in examining relationships between variables as compared to a researcher who is testing a nomological net, i.e., theory, may identify different error rate units for the identical set of statistical tests.

As previously stated, the researchers under Scenario 1 considered the two dependent variables to be part of one overall analysis. Thus, they may identify only one error rate unit for the study. The researchers in Scenario 2, however, considered the two dependent variables as constituting two separate analyses. Under this scenario the researchers may identify two error rate units. Similar to the researchers in Scenario 2, the researchers in Scenario 3 viewed the study as containing two error rate units with each unit based on the analysis of each dependent variable.

Step 2: Identify the Number and the Nature of the Statistical Tests in Each Error Rate Unit.

Researchers can determine the alpha level for each statistical test in a given error rate unit if they know the following two pieces of information: (a) the overall alpha level for the unit, and (b) the number of exploratory statistical tests in the unit. Before researchers can obtain this second piece of information, they must determine the nature of the hypothesis being tested by each statistical test. That is, the researchers must be able to stipulate whether a given statistical test is based on theory and/or previous research findings or whether the test is exploratory in nature. As previously stated, we refer to the statistical tests of hypotheses based on theory and/or previous research results as theory-based tests. In addition, we refer to the statistical tests of hypotheses for which the researchers do not predict the results as exploratory tests.

As previously mentioned, the researchers under Scenarios 1 and 2 approached the study in an exploratory manner. Thus, under these scenarios, each statistical test would be considered as an exploratory test. Unlike the researchers in Scenarios 1 and 2, the researchers in Scenario 3 predicted that the interaction effects between the treatments and gender would exist for both of the dependent variables. Thus, the statistical tests for the interaction effects were considered to be theory-based tests.

Once the researchers classify the statistical test of each hypothesis as either theory or exploratory based, they must identify the number of exploratory statistical tests in each error rate unit. To illustrate this point, consider the three research scenarios for the hypothetical teacher efficacy study. In Scenario 1 the researchers formed one error rate unit that contained the four tests of the main effects and the two tests of the interaction effects. Since each of these six tests was classified as exploratory, the error rate unit contained a total of six tests that would impact the alpha adjustment process. Under Scenario 2, each of the two error rate units contained two main effects tests and one interaction effect test. Since each of these three tests was classified as exploratory, each unit contained three tests that would affect the alpha adjustment process. Under Scenario 3, which also contained two error rate units, the researchers classified the two interaction effects as theory based. Thus, the alpha levels corresponding to the statistical tests of the two interaction effects would not be adjusted or affect any subsequent alpha adjustment process. If the interaction effect in either of the error rate units was found not to be statistically significant, the researchers would continue their investigation by testing the two main effects with tests that would be considered as exploratory in nature. Thus, each error rate unit would contain two exploratory tests. These two exploratory tests would impact the alpha adjustment process implemented for any further statistical testing.

To summarize, Scenario 1 contains six exploratory statistical tests. Scenario 2 has three exploratory statistical tests in each of its two error rate units. In Scenario 3 the alpha level for the statistical test of the interaction effect in each of the two error rate units would not be adjusted. If the statistical test of the interaction effect in a given error rate unit is not significant, the unit would contain two additional tests, which would be exploratory in nature.

Step 3: Adjusting the Alpha Levels of the Various Statistical Tests.

Once the number of exploratory based statistical tests contained in the error rate units has been identified and the overall alpha level has been set, the alpha level for each statistical test can be adjusted. Researchers have numerous adjustment methods at their disposal. We suggest, however, that researchers strongly consider implementing a Bonferroni-type adjustment procedure for two reasons. First, a Bonferroni-type adjustment tends to be robust with respect to violations of homogeneity of variance (Kromrey & La Rocca, 1995). Second, Bonferroni-type adjustments afford researchers greater flexibility in a number of ways, which include the following: (a) controlling Type I error rates in studies that contain multiple statistical tests of correlation values and regression coefficients and (b) adjusting alpha levels in a manner which reflect a concern that the consequences of committing Type I errors are not equally serious for all of the statistical tests (see Kirk, 1994, p. 97 for an illustration of this unequal weighting scheme).

One concern a researcher may have when using a Bonferroni-type adjustment procedure is its lack of power as compared to the power levels produced by other adjustment procedures. We believe the importance of this concern can be significantly diminished in two ways. First, if researchers take the position that the alpha levels will only be adjusted based on the number of exploratory hypotheses being tested in the error rate unit, the number used to adjust the alpha levels of the statistical tests will often be decreased. This practice will increase the power levels of the tests and, thus, diminish the differences in power levels between the Bonferroni-type adjustment procedures and other adjustment procedures. As noted by Kromrey and La Rocca (1995) in their study on the relative power levels of various Type I error rate adjustment procedures in ANOVA studies, "Researchers who are interested in only a smaller number of contrasts among cell means will find that the Dunn procedure and its modifications [Bonferroni-type adjustment procedures] provide more power than was obtained in this research" (p. 360).

If a study is exploratory in nature, that is, the researchers are not willing to predict the outcomes of their statistical tests based on theory and the results of previous research, we take the position that the Bonferroni-type adjustment procedures appropriately diminish the power levels of the tests. In such studies, we believe it should be more difficult to produce statistically significant findings than in studies that are not exploratory.

Second, the potential differences between the power levels of Bonferroni-type adjustment procedures and other adjustment procedures, which are often more complex and less flexible, can be reduced by utilizing a step-down Bonferroni-type adjustments such as those proposed by Holm (1979) and Shaffer (1986). Although these step-down Bonferroni-type adjustments will not be presented in this article, we encourage interested researchers to consider their use when the power levels are a significant concern.

To illustrate how the third step in the adjustment procedure can be implemented, assume that the researchers want to maintain the overall alpha levels at .05 for each error rate unit. As previously noted, the researchers identified only one error rate unit in Scenario 1, which contained both dependent variables. This unit contained a total of six statistical tests, i.e., three tests for each dependent variable. The factors tested for each dependent variable were the following: (a) the interaction effect, (b) the treatment main effect, and (c) the gender main effect. In addition to these six tests, if either or both of the treatment main effects tests were significant, pairwise comparison tests of the treatment means would also be conducted. All of the statistical tests under this scenario were considered to be exploratory. Thus, the alpha levels for the main effects and the interaction effects statistical tests would be set at .0083, which is equal to the overall alpha level of the error rate unit (.05) divided by the number of exploratory statistical tests contained in the error rate unit (6). If either of the treatment main effects statistical test was found to be significant, the alpha level for each pairwise comparison test would be set at .0027, which is equal to the adjusted alpha level of the treatment main effect statistical test (.0083) divided by the number of pairwise comparisons (3). See Table 1 for a list of the adjusted alpha levels of the statistical tests contained in Scenario 1.

As previously noted, the two error rate units identified in Scenario 2 were based on the two dependent variables, i.e., the two types of efficacy scores. Thus, each unit contained one statistical test of the interaction effect and one statistical test of each of the two main effects. In addition to these three tests, if either or both of the treatment main effects tests were significant, pairwise comparison tests of the treatment means would also be conducted. Similar to Scenario 1, it is assumed that each test included in Scenario 2 is considered to be exploratory. Thus, the alpha level used to statistically test the interaction effect and each main effect contained in each error rate unit is .017, which is obtained by dividing the overall alpha level of the error rate unit (.05) by the number of statistical tests of exploratory hypotheses in the error rate unit (3). The alpha level for each pairwise comparison test would be .0057, which is calculated by dividing the adjusted alpha level for the treatment main effect (.017) by the number of pairwise comparisons (3). See Table 1 for a list of the adjusted alpha levels of the statistical tests contained in Scenario 2.

The researchers in Scenario 3 identified the same two error rate units as those identified by the researchers in Scenario 2. Unlike Scenario 2, however, each error rate unit contained one theory-based statistical test of the interaction effect. Thus, the alpha level used to statistically test each of the interaction effects was not adjusted, i.e., it was set at .05. If a given interaction effect was not significant, however, the researchers would consider the tests of the corresponding treatment and gender main effects as exploratory. The alpha levels for the corresponding main effects tests would be adjusted. The adjusted alpha level for each of those two main effect tests would be .025, which is equal to the overall alpha level of the error rate unit (.05) divided by the number of exploratory tests in the unit (2). If a given treatment effect is significant,

Table 1
Adjustments to the Alpha Levels of the Statistical Tests Continued In Scenarios 1, 2, and 3

		Adjusted Alpha Levels		
Dependent Variable	Variables Tested	Scenario 1	Scenario 2	Scenario 3
Teaching Efficacy	Treatment	.0083	.0170	.0250
	A vs B[1]	.0028	.0057	.0083
	A vs C[1]	.0028	.0057	.0083
	B vs C[1]	.0028	.0057	.0083
	Gender	.0083	.0170	.0250
	Interaction Effect	.0083	.0170	.0500
Personal Efficacy	Treatment	.0083	.0170	.0250
	A vs B[1]	.0028	.0057	.0083
	A vs C[1]	.0028	.0057	.0083
	B vs C[1]	.0028	.0057	.0083
	Gender	.0083	.0170	.0250

the alpha level for each of the three pairwise comparison tests of the treatment means would be set at .0083, which is equal the adjusted alpha level for the treatment main effect (.025) divided by the number of pairwise comparison tests (3). See Table 1 for a list of the adjusted alpha levels of the statistical tests contained in Scenario 3.

Comparison of the Alpha Levels in the Three Scenarios

When one compares the alpha levels for the tests conducted under each scenario, which are listed in Table 1, a number of points should be noted. First, the numerous tests of exploratory hypotheses contained in Scenario 1 cause their alpha levels to be quite restrictive. As previously discussed, we believe these stringent alpha levels are appropriate in a study that is exploratory in nature. Second, the identification of two error rate units in Scenario 2 produced higher alpha levels than the ones calculated in Scenario 1, which contained only one error rate unit. The formation of one or two error rate units can be debated. The point is, however, the use of this three-step adjustment procedure will force the study's researchers to reflect on this decision before adjustments in the alpha levels of the various test are made. Third, the clear identification of the error rate units by the researchers and, hopefully, a discussion of the rationale used to form the units, will allow the readers of their studies to have a more complete understanding of the data analysis procedures.

Summary

To protect against inflated Type I error rates in studies that contain multiple statistical tests, researchers need to consider implementing an adjustment procedure. The three-step adjustment procedure we have presented first requires the researchers to identify the error rate units that will serve as the basis for the adjustment process. Next, they must identify the number and nature of the statistical tests contained in each error rate unit. Finally, they must adjust the alpha levels of some, if not all, of the individual statistical tests contained in the error rate unit using a Bonferroni-type adjustment procedure.

This three-step adjustment procedure provides researchers with a tool that is robust, flexible, and easy to apply. More important, this procedure requires researchers to reflect on the selection of the conceptual error rate unit on which the Type I error adjustments are based. We believe that such reflection will lead to better research and program evaluation.

In addition to encouraging the use of this three-step adjustment approach, we believe that it is important for educational researchers to engage in philosophical discussions and research to identify the most appropriate error rate units for specific types of research questions and situations. We believe this type of investigation may prove to be just as valuable, if not more so, for the fields of educational research and program evaluation than additional Monte Carlo and analytical studies on Type I error rate correction procedures.

References

Brown, M.B., & Forsythe, A.B. (1974). The ANOVA and multiple comparisons for data with heterogeneous variance. *Biometrics, 30*, 719-724.

Duncan, D.B. (1955). Multiple range and multiple F tests. *Biometrics, 11*, 1-42.

Dunn, O.J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56*, 52-64.

Fisher, R.A. (1949). *The design of experiments*. Edinburgh: Oliver and Boyd Ltd.

Hays, W.L. (1988). *Statistics* (4th ed.). Fort Worth, TX: Holt, Rinehart and Winston.

Hockberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*, 800-802.

Holland, B.S., & Copenhauer, M.D. (1987). An improved sequentially rejective Bonferroni Test procedure. *Biometrics, 43*, 417-423.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.

Kirk, R.E. (1994). Choosing a multiple comparison procedure. In B. Thompson (Ed.), *Advances in Social Science Methodology* (vol 3, pp. 77-121). Greenwich, CT: JAI Press.

Kirk, R.E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks/Cole.

Kramer, C.Y. (1956). Extension of multiple range test to group means with unequal numbers of replications. *Biometrics, 12*, 307-310.

Kromrey, J.D., & La Rocca, M.A. (1996). Power and type I error rates of new pairwise multiple comparison procedures under heterogeneous variances. *The Journal of Experimental Education, 63*(4), 343-362.

Kuels, M. (1952). The use of studentized range in connection with an analysis of variance. *Euphytica, 1*, 112-122.

McHugh, R.B., & Ellis, D.S. (1955). The "postmortem" testing of experimental comparisons. *Psychological Bulletin, 52*, 425-428.

Newman, D. (1939). The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika, 31*, 20-30.

Newman, I., & Fry, J. (1972). A note on multiple comparisons and a comment on shrinkage. *Multiple Linear Regression Viewpoints, 2*, (3), 36-39.

Ryan, T.A. (1959). Multiple comparisons in psychological research. *Psychological Bulletin, 56*, 26-47.

Ryan, T.A. (1962). The experiment as the unit for computing rates of error. *Psychological Bulletin, 59*, 301-

305.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika, 40*, 87-104.

Shaffer, J.P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association, 81*, 826-831.

Spjøtvoll, E., & Stoline, M.R. (1973). An extension of the T-method of multiple comparisons to include the cases with unequal sample signs. *Journal of the American Statistical Association, 68*, 975-978.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Toothaker, L.E. (1991). *Multiple comparisons for researchers*. Newbury Park, CA: Sage.

Tukey, J.W. (1953). *The problem of multiple comparisons*. Ditto: Princeton University.

Wilson, W. (1962). A note on the inconsistency inherent in the necessity to perform multiple comparisons. *Psychological Bulletin, 59*, 296-300.

Winer, B.J., Brown, D.R., & Michols, K.M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Isadore Newman is a Distinguished Professor at the University of Akron.

John Fraas is a Trustees' Professor at Ashland University.

John M. Laux is a doctoral student in the Counseling Psychology Department at the University of Akron.