DOCUMENT RESUME

ED 449 216 TM 032 361

AUTHOR	Baugh, Frank G.
TITLE	Increasing the Generalizability of ANOVA Results by Judicious Selection of Fixed-, Random-, and Mixed-Effects ANOVA Models.
PUB DATE	2001-02-00
NOTE	24p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Analysis of Variance; Research Methodology
IDENTIFIERS	*Fixed Effects; *Random Effects

ABSTRACT

The analysis of variance (ANOVA) is a frequently used statistical procedure by which the equality of more than two population means can be tested without inflating the Type I error rate (D. Hinkle, W. Wiersma, and S. Jurs, 1998). Fixed-, random-, and mixed-effects ANOVA models are each capable of yielding interesting and useful results when applied in appropriate situations. However, the random- and mixed-effects models offer the added benefit of increasing the generalizability of results. This paper illustrates "rules of thumb" (C. Hicks, 1973) researchers can use to test all three models, explores the factors that should bear on model selection (B. Frederick, 1999), and explains how the Statistical Package for the Social Sciences (SPSS) can be use to evaluate all three models. An appendix contains the SPSS program to analyze fixed- and random-effects models. (Contains 3 tables and 19 references.) (Author/SLD)

Running head: INCREASING THE GENERALIZABILITY OF ANOVA

Increasing the Generalizability of ANOVA Results by Judicious Selection of Fixed-, Random-, and Mixed-Effects ANOVA Models

Frank G. Baugh

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 1-3, 2001

Abstract

The analysis of variance (ANOVA) is a frequently utilized statistical procedure whereby the equality of more than two population means can be tested without inflating the Type I error rate (Hinkle, Wiersma & Jurs, 1998). Fixed-, random-, and mixed-effects ANOVA models are each capable of yielding interesting and useful results when applied in appropriate situations. However, the random- and mixed-effects models offer the added benefit of increasing the generalizability of results. This paper illustrates "rules of thumb" (Hicks, 1973) researchers can use to test all three models, explores what factors should bear upon model selection (Frederick, 1999), and explains how SPSS can be used to evaluate all three models.

Increasing the Generalizability of ANOVA Results by Judicious Selection of Fixed-, Random-, and Mixed-Effects ANOVA Models

Generalizability of results is often a common goal among researchers working on a variety of research questions. Indeed, the area of inferential statistics was developed with this goal in mind. Often collecting data from every member of a population of interest is not feasible. As a result methods have been developed for drawing a sample or smaller subset of the population, collecting data from this sample, and analyzing the data in such a way that conclusions can be drawn about the larger population. Without the ability to generalize results beyond the present sample to a larger population, many studies would fail to produce useful or meaningful conclusions.

It is a common and well-known practice among researchers to use samples of people, presuming that these samples are representative of a larger population, and thereby generalize their results beyond the sample data. Less common but also effective in increasing the generalizability of results is the sampling of categories of independent variables in service of generalizing to the population of all categories defining a variable. In other words, just as people can be sampled, the levels in an ANOVA way (factor) can also be randomly selected, and then results can be generalized beyond the levels used to all the possible levels. For example, if therapy sessions reasonably could be any time between 45 and 90 minutes in length, a researcher could randomly select only three of the 46 possible levels (e.g., 46, 52, 88) in this "time" way, and yet still generalize findings to all 46 times.

As all researchers are aware, ANOVA is a frequently utilized statistical procedure whereby one is able to test the equality of more than two population means without inflating the Type I error rate (Hinkle, Wiersma & Jurs, 1998). Three ANOVA models exist – fixed-effects (Model I), random-effects (Model II), and mixed-effects (Model III). If the ways in the design (a) include all the possible levels of that way or (b) do not include all the possible levels but still are the only levels of interest, the design is a "fixed" model (Jackson & Brasher, 1994). For example, if a researcher uses male and female as levels in a gender way, the design is inherently fixed. Alternatively, if the researcher in the therapy study chooses time levels of 45, 60, and 90, and did so as a choice and not through random sampling, and does not generalize to other times, the design is fixed. Conversely, a random-effects model utilizes a way(s) with randomly selected levels from among all the possible levels in that way. A mixed-effects model occurs only in multi-way designs in which some ways are fixed and some are random (Jackson & Brasher, 1994). For example, a "gender" way (male, female) by randomly selected "therapy durations" way (46, 51, 88 minutes) study would invoke a "mixed" model.

Although there are three ANOVA models, the fixed-effects model is most commonly used. In fact, most statistical computing packages presume as a default a fixed-effects model. As a result, many researchers unknowingly choose a fixed-effects model when a random- or mixed-effects model may be more appropriate. However, informed researchers can take the computer results and fairly easily re-calculate a portion of results to obtain correct values for their actual models.

This paper will illustrate "rules of thumb" (Hicks, 1973) that researchers can use to test all three models. Additionally, factors that should bear upon model selection (Frederick, 1999) will also be explored. Finally, use of SPSS to evaluate all three models will be illustrated.

Models

Fixed-effects models differ from random-effects model in the levels of factors included in the models and the resulting inferences. Because fixed factors consist of all possible levels or of specifically chosen levels, inferences can only be drawn about the levels used, neglecting all other levels that could possibly be used. These levels themselves represent all the interesting conditions. In this manner only the levels that are of interest to the researcher are included in the analysis (Hays, 1981).

In contrast, random ways consist of levels that are a sample of a larger population of levels. The researcher in this case has no specific interest in the levels used in the study, but intends to make inferences about all the levels that could have possibly been included (Hays, 1981; Jackson & Brasher, 1994; Ostle & Malone, 1988).

The fixed-effects model considers a different null hypothesis than does the random-effects model. Because the researcher is only interested in the specific levels used in the study, the null hypothesis that there is no difference among the effects refers only to the levels included in the study. On the other hand, the null hypothesis of no difference in the random-effects model refers to the entire population of levels of which the levels in the study are only an arbitrary subset (Ostle & Malone, 1988).

Hays (1981) indicated that in social science research the mixed-effects model is perhaps more common in multi-way designs than that of the random model. Researchers utilizing the mixed-effects model have at least one factor with all levels of interest included in the study and at least one factor in which the population of possible levels is of interest but only a subset of possible levels is actually used in the study (Hays, 1981).

It is important to note that the same way considered fixed in one study could be considered random in another study. Coleman (1979), commenting on the classification of ways, stated, "…there are usually several correct ways to analyze an experiment, and … the better choice is more a matter of wisdom than mathematical correctness" (p. 243). The consequences of classifying a factor as fixed when it should be classified as random has, however, been outline by a number of writers (Clark, 1973; Forster & Dickinson, 1976; Jackson & Brasher, 1994; Santa, Miller, & Shaw, 1979; Wickens & Keppel, 1983). Clark (1973), in his article "Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research," pointed out that researchers forego any statistical evidence of generalizability of their results when they treat words as fixed-effects and then present conclusions suggesting that their findings generalize beyond these selected words to a larger population of words.

Jackson and Brasher (1994) contended that limited or lessened generalizability results from an analysis wrongly classifying random-factors as fixed. In fact, subtle statistical errors occur in this situation resulting from a shift regarding which null hypothesis is actually being tested. The consequence is not an invalid test, but a test that is valid for a hypothesis, which the researcher did not intend to test. Any conclusions the researcher makes about the original research question will not be supported by the statistical results (Jackson & Brasher, 1994). Richter and Seay (1987) provide convincing evidence of the consequences of the misclassification of ways. Their re-analysis using a random-effects model of an earlier experiment that yielded seven statistically significant results using a fixed-effects model, revealed that only one of the original seven results remained statistically significant!

Obviously, sometimes using mixed- or random-effects models can have important benefits of yielding more generalizable results. For example, Hays (1981) noted that "the random-effects models is designed especially for experiments in which inferences are to be drawn about an entire set of distinct treatments or factor levels, including some not actually observed" (p. 376). The design can be very efficient and economical. Indeed, it has even been argued that in some cases the use of fixed-effects models makes little sense (cf. Clark, 1973).

The question, then, of how to classify a way becomes of utmost important to the researcher. Unfortunately, a debate exists as to which factors should be classified as fixed and which should be classified as random. All will agree a factor is fixed when all of its possible levels are included in a study. Nevertheless, controversy arises about how to classify a factor when it is not feasible or desirable to include all the levels of the factor in the study. Some argue that ways should be classified as random any time the levels of the way included in the study do not exhaust the population of acceptable levels that could have been included, even if the levels of the way are not chosen through a random process (Clark, 1973; Richter & Seay, 1987). Researchers should make every effort to choose levels included in their study through a non-arbitrary, random process. Yet, if random sampling is not possible, the identification of the specific population to which their results generalize becomes the only difficulty (Clark, 1973). Kennedy and Bush (1985) concurred noting that the classification of variables as random and the subsequent application of the random-effects model should be expanded to include those variables whose levels appear to be representative of the larger population regardless of whether random procedures of selection were employed.

Conversely, Wike and Church (1976) challenged this position indicating it contradicts traditional criteria used by statisticians of classifying ways based on the number of levels intended to be included in the experiment and the procedure employed to identify those levels from among all possible levels. Whenever p, the levels included in the experiment, equals P, the possible levels that could have been included, the factor is fixed. Also, if p is chosen through any process not deemed random and is less then P, the factor is still consider fixed. Some have argued the factor is classified as random only if the p levels are not equal to P and are selected by randomization (Wike & Church, 1976).

Richter and Seay (1987) agreed that treating a way as random when levels not included and not randomly sampled could compromise the accuracy of the ANOVA. However, the writers shrewdly pointed out that participants are regularly treated as a random way even when they are not sampled in a truly random manner. According to Clark (1973), failing to classify ways as random when all possible levels are not included sacrifices generality across levels. Others (Richter & Seay, 1987; Santa, Miller, & Shaw, 1979) also indicate that evidence of generality of an effect across other levels requires ways be treated as random in the ANOVA model (Richter & Seay, 1987). Coleman (1963) piercingly illustrates the consequence of not utilizing the random-effects model in language studies,

> Many studies of verbal behavior have little scientific point if their conclusions have to be restricted to the specific language materials that were used in the experiment. It has not been customary, however, to perform significance tests that permit generalization beyond these specific materials, and thus there is little statistical evidence that such studies could be successfully replicated if a different sample of language materials were used. (p. 219)

However, Wike and Church (1976) disputed the claim that generalizability depends on the random selection of levels. A randomization process for including levels may be inefficient as well as unsuccessful in sufficiently representing the population of levels that could be selected. It is entirely possible that levels selected could be too similar to yield the differences present in the entire population causing inflation of the Type II error rate (failing to reject the null when it should be rejected). A controlled choice of representative levels for inclusion can often prove more desirable than a random process of selection (Wike & Church, 1976). However, Coleman (1979) indicates that failure to use tests of generalization can lead to both Type II and Type I errors.

Analytic Criteria

The lack of agreement about how to classify ways when using the ANOVA statistic can be confusing to the researcher. But it is suggested that when possible, utilize a random-effects model adhering to traditional criteria. For example, Hicks (1973) noted that,

> It is not reasonable to decide after the data have been collected whether the levels are to be considered fixed or random. This decision must be made prior to the running of the experiment, and if random levels are to be used, they must be chosen from all possible levels by a random process. (p. 173)

So what factors should bear upon model selection? Several writers have delineated guidelines that apply to the issue of selection (Frederick, 1999; Hays, 1981; Jackson & Brasher, 1994; Longford, 1993; Shavelson & Webb, 1991). The first factor to consider is related to the substitutive nature of the levels. Treating a way as random is preferable when all of the possible levels of the way are equally analogous. In other words, if any of the levels could be substituted in the design without changing the original research question, then the way should be considered random. The context of the experiment, therefore, dictates classification of the way (Longford, 1993; Shavelson & Webb, 1991).

Consider the following example in which a researcher wishes to examine the effectiveness of three methods of psychotherapy (cognitive therapy, behavior therapy, person-centered therapy) in treating depression with particular attention to the number of sessions. Assuming the number of sessions could reasonably be anywhere within the range of three to eighteen, the researcher could randomly select a subset of four different session numbers from the total population of possible sessions (3 to 18). The researcher could then generalize results of the experiment beyond the numbers of session included in the analysis to the entire continuum of session numbers. It is inconsequential whether the specific session numbers included in the study were 3, 9, 13, and 17 or 4, 10, 14, and 18 – the specific number of sessions are of no particular interest to the researcher. Substituting one treatment length for another does not alter the experiment. On the contrary, if different methods of psychotherapy were substituted, the experiment would have a modified meaning.

The interpretations a researcher hopes to reach based on the results of an investigation forms the foundation of the second method of determining whether a way should be classified random or fixed (Jackson & Brasher, 1994). Fixed and random ways are distinguishable by the inferences that can be reasonably drawn from them. A researcher who wishes to draw conclusions only about the specific levels included in the investigation would be obliged to classify the way as fixed. The way must be classified as random if the researcher seeks findings that will allow for extrapolation beyond the levels represented (Hays, 1981; Jackson & Brasher, 1994). The psychotherapy example illustrates this difference. Consider the generalizations the researcher can make if the "session" way is treated as fixed. It could be reasonably concluded, for example, that cognitive therapy is more effective in treating depression than person-centered or behavioral therapy after 4, 11, and 14 sessions provided these were the therapy durations represented. Alternatively, treating the "session" way as random allows the researcher to reasonably infer from the results that cognitive therapy is more effective than behavioral or person-centered therapy in treating depression across all therapy durations.

A final method for making the distinction is set forth by Jackson and Brasher (1994). The writers recommend researchers consider the answer to the question, will generalizations only to the specific levels of the way provide any useful or notable information? An affirmative answer to the question demonstrates cause to classify the way as fixed, while the realization that inferring only to the represented levels reveals not very useful information dictates random classification of the way (Jackson & Brasher, 1994). For example, identifying cognitive therapy as the more effective treatment of depression of the three possible psychotherapies across all durations ranging from 3 to 18 sessions is more meaningful than identifying cognitive therapy as more effective after 4, 11, and 14 sessions.

Assumptions

The ANOVA model is based on several assumptions. Although related, the assumptions accompanying random- and mixed-effects models differ from those of the fixed-effects model (Jackson & Brasher, 1994). Assumptions of the random- and mixed-effects models have been outlined and discussed by several writers (Jackson & Brasher, 1994; Hays, 1981; Hicks, 1973; Mason, Gunst, & Hess, 1989; Ott, 1984). Mason, Gunst, and Hess (1989) concisely summarized these: (1) each level of the random ways present in the study are a random subset of the population of levels that are of interest, (2) main effects and interactions represented by random ways are presumed statistically independent, (3) experimental errors are statistically independent, (4) a normal distribution is presumed. In addition, the homogeneity of variance assumption applies along with the concept of sphericity (Jackson & Brasher, 1994). Because the violation of assumptions negatively impacts the accuracy of and hence the degree of confidence one should place in the results of the ANOVA model, researchers should not only be aware of such violations but also anticipate their occurrence (Jackson & Brasher, 1994).

Determining the Correct Error Term: "Rules of Thumb"

The presence of randomized levels in the random- and mixed-effects models necessitates slightly different procedures for calculating the *F* ratio test statistic. It is important to note *the procedures for determining the sum of squares, degrees of freedom, and the mean squares are the same across all three models*. However, determining *F* (calculated) requires an adjustment in the error term (denominator) of the equation across the three models.

In the fixed-effects model, the error term is always the Mean Square Error (Hinkle, Wiersma, & Jurs, 1999). However, variance attributed to the random sampling of levels requires the error term for random ways include other essential elements. With the additional component in the denominator, all sources of variance can be removed through division (Kennedy & Bush, 1995).

Jackson and Brasher (1994) noted that the appropriate test statistic for various research designs are unique to the design. Therefore, it is necessary for researchers to be familiar with procedures for determining the test statistic that fits their specific design

(Jackson & Brasher, 1994). Utilizing the "Expected Mean Square" [E(MS)] researchers can estimate the appropriate error terms and hence choose the correct statistic (Hinkle, Wiersma, & Jurs, 1999; Jackson & Brasher, 1994; Mason, Gunst, & Hess, 1989). According to Jackson and Brasher (1994), the E(MS) is the average value of variances if an experiment was conducted infinitely many times maintaining identical structure. Unfortunately, the calculations required to compute E(MS) are complex and tedious (Hicks, 1973). Simple steps, however, can be employed to estimate E(MS) fairly accurately (Hicks, 1973; Ott, 1984).

Returning to our psychotherapy example will allow a detailed application of the steps outlined by Hicks (1973) and Ott (1984). Assume the researcher in the example utilized a balanced, 3 x 4 design in which the A way (types of therapy) is fixed and the B way (number of sessions) is random. For the sake of ease, we will use a small data set of 24 observations, two in each cell.

1. The first step in the process of constructing the two way table is to list the ways in the model as row labels.

$\mathbf{A}_i$
$\mathbf{B}_j$
$\mathbf{AB}_{ij}$
$\mathrm{E}_{k(ij)}$

2. Record the subscripts in the model as column headings. Also, write over the subscripts an F if the levels of that way are fixed, or R if they are random. Interactions

involving a random way are considered random. Above the letters, record the number of observations corresponding to each term (This is the number of levels for the main effects and the number of participants in each cell for the error term).

	4 F i	3 R j	2 R k
$\mathbf{A}_i$			
$\mathbf{B}_j$			
$\mathbf{AB}_{ij}$			
$E_{k(ij)}$			

3. In each column copy the top number representing observations to every row in the column that does not contain the subscript of the column.

	4 F i	3 R j	2 R k
$\mathbf{A}_i$		3	2
$\mathbf{B}_j$	4		2
$\mathbf{AB}_{ij}$			2
$E_{k(ij)}$			

4. Under any column containing a subscript that appears in parentheses, record a one in the error row label.

	4 F I	3 R j	2 R k
$\mathbf{A}_i$		3	2
$\mathbf{B}_j$	4		2
$\mathbf{AB}_{ij}$			2
$E_{k(ij)}$	1	1	

5. Finally, place a zero in any empty space in columns representing a fixed way, and a one under those representing a random way.

	4 F I	3 R j	2 R k
$\mathbf{A}_i$	1	3	2
$\mathbf{B}_j$	4	0	2
$\mathbf{AB}_{ij}$	1	0	2
$E_{k(ij)}$	1	1	1

The completed table can now be used to estimate the E(MS) for each term in the model. To do this, cover the column(s) that contain a subscript which matches the subscript of the row label. For example, to find the E(MS) for the $\mathbf{A}_i$ term cover column i. For the $\mathbf{AB}_{ij}$ interaction term, you would cover columns i and j. The key to identifying which rows will contribute to a specific way's equation are the subscripts. Any row that shares a common subscript with the particular way will contribute to E(MS) equation of that way.

For example, the $\mathbf{A}_i$ way E(MS) equation will include variance terms of the $\mathbf{A}_i$ way effect, $\mathbf{AB}_{ij}$ interaction, and error term because the *i* subscript is present in each of the rows. However, the $\mathbf{B}_j$ way effect does <u>not</u> contribute to this equation because the *i* subscript is <u>not</u> present in this row. After determining which effects to include in each row's equation, multiply the numbers that are not covered in each row by that's rows variance term. This term will be ϕ for the fixed ways and σ^2 for the random ways. Combining the appropriate variances through the additive process forms the appropriate E(MS) equation for the particular row effect. The A way's E(MS) equation, for example, would be:

(3)(2)(ϕ) from the A way + (2)(σ^2) for the AB interaction + (1 x 1) (σ^2)

$$= 6\phi_A + 2\sigma_{AB}^2 + \sigma_E^2$$

	4 F i	3 j	2 R k	E(MS)
A	1	3	2	$6\phi_A + 2\sigma_{AB}^2 + \sigma_E^2$
B	4	0	2	$8\sigma_B^2 + \sigma_E^2$
AB	1	0	2	$2\sigma_{AB}^2 + \sigma_E^2$
E	1	1	1	σ_E^2

By following the "rules of thumb" one can quickly compute E(MS) equations which account for all of the variance (within error and sampling error) present in a given way. Remember the purpose of deriving the E(MS) is to determine the appropriate denominator for the *F* ratio calculations. Going back to our example, consider the A way E(MS) equation:

$$6\phi_A + 2\sigma_{AB}^2 + \sigma_E^2 .$$

The first and second term in the equation represents the variance from the A way main effect and the AB interaction, respectively. The last term is the error variance term.

The goal of the *F* ratio in this situation, then, would be to separate out the variance that is <u>only</u> associated with the A way effect. As you can see in the equation, the additional variance comes from the AB interaction and the error effects. Therefore, the *F* ratio for the A way would be computed using as a denominator <u>both</u> the mean square interaction and the mean square error. The *F* ratios for the other effects would be:

B way main effect = $8\sigma_B^2 + \sigma_E^2 / \sigma_E^2$, and

AB interaction effect = $2\sigma_{AB}^2 + \sigma_E^2 / \sigma_E^2$.

SPSS Example

The advancement of technology has led to the widespread use of computer packages such as SPSS in analyzing data. An understanding of the previous discussion regarding the differences in models and methods for determining which model to utilize is necessary for researchers who wish to take advantage of increased generality of random-effects ANOVA results. In addition, it is helpful for researchers to also possess familiarity with the procedures for analyzing the three models in SPSS. Therefore, a small heuristic data set in the context of our psychotherapy example will be utilized to make these procedures accessible.

Table 1 presents the hypothetical data for 24 participants in the 3 x 4 design with 2 observations per cell. The A way is the types of therapy (cognitive therapy, behavioral therapy, person-centered therapy) and the B way is the number of sessions (4, 7, 12, 17). The dependent variable for this example are scores on the Beck Depression Inventory (BDI). Appendix A presents the SPSS commands to perform the UNIANOVA with both ways fixed and then with both ways random.

Table 2 presents the results of the analysis with both ways fixed. Table 3 presents the results with both ways random. Note the only differences in the results between the tables are the *F* (critical) values and the level of statistical significance for treatments and sessions. This is because different denominators are used to compute the *F*s in the two models, as explained in the table notes. SPSS Version 10.0 is unable to correctly analyze the mixed-effects model. In attempting to analyze the "mixed" model it divides by the wrong error term and thus distorts the results.

Discussion

Fixed-, random-, and mixed-effects ANOVA models are each capable of yielding interesting and useful results when applied in appropriate situations. As is the case with all areas of inferential statistics, the decision regarding which model to utilize is often less an issue of right or wrong and depends on the researcher's intended purpose (Coleman, 1979). Researchers will do well to give ample thought for guidance with model selection to the levels included in their study and the conclusions they wish to draw from their results. The random- and mixed-effects ANOVA models offer the added benefit of increasing the generalizability of results. Unfortunately, many researchers fail to make use of these models in favor of the fixed model. One possible explanation for this sometimes costly decision is the reality that most computer packages for behavioral science statistical procedures assume a fixed model. However, informed researchers can follow the procedures outlined in this paper in order to obtain corrected results and thereby increase the generalizability of their results.

References

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 12, 335-359.

Coleman, E.B. (1964). Generalizing to a language population. Psychological Reports, 14, 219-226.

Coleman, E.B. (1979). Generalization effects vs random effects: Is $\sigma_{TL}{}^{2}$ a source of Type 1 or Type 2 error. Journal of Verbal Learning and Verbal Behavior, 18, 243-256.

Forster, K.I., & Dickinson, R.G. (1976). "More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F1, F2, F', and min F. Journal of Verbal Learning and Verbal Behavior, 15, 132-142.

Frederick, B.N. (1999). Fixed-, random-, and mixed-effects ANOVA models: A user-friendly guide for increasing the generalizability of ANOVA results. In B. Thompson, Advances in social science methodology (Vol. 5, pp. 111-122). Stamford, CT: JAI Press.

Hays, W.L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart, and Winston.

Hinkle, D.E., Wiersma, W. & Jurs, S. G. (1998). Applied statistics for the behavioral sciences (4th ed.). Boston: Houghton Mifflin Company.

Hicks, C.R. (1973). Fundamental concepts in the design of experiments. New York: Holt, Rinehart, and Winston.

Jackson, S. & Brashers, D.E. (1994). Random factors in ANOVA. Thousand Oaks, CA: Sage.

Kennedy, J.J. & Bush, A.J. (1985). An introduction to the design and analysis of experiments in behavioral research. New York: University Press of America.

Longford, N.T. (1993). Random coefficient models. New York: Oxford University Press.

Mason, R.L., Gunst, R.F., & Hess, J.L. (1989). Statistical design and analysis of experiments. New York: Wiley & Sons.

Ott, L. (1984). An introduction to statistical methods and data analysis (2^{nd} ed.). Boston: Prindle, Weber, & Schmidt.

Ostle, B. & Malone, L.C. (1988). Statistics in research: Basic concepts and techniques for research workers (4^{th} ed.). Ames, IA: Iowa State University Press.

Richter, M.L. & Seay, M.B. (1987). Anova designs with subjects and stimuli as random effects: Applications to prototype effects on recognition memory. Journal of Personality and Social Psychology, 53, 470-480.

Santa, J.L., Miller, J.J., & Shaw, M.L. (1979). Using quasi *F* to prevent alpha inflation due to stimulus variation. Psychological Bulletin, 86, 3-46.

Shavelson, R.J. & Webb, N.M. (1991). Generalizability theory: a primer. Newbury Park, CA: Sage.

Wickens, T.D., & Keppel, G. (1983). On the choice of design and of test statistic in the analysis of experiments with sampled materials. Journal of Verbal Learning and Verbal Behavior, 22, 296-309.

Wike, E.L. & Church, J.D. (1976). Comments on Clark's "The language-as-fixed-effect fallacy". Journal of Verbal Learning and Verbal Behavior, 15, 249-255.

Table 1
Heuristic Data

Tx	Sessions	BDI
1	1	20
1	1	19
1	2	18
1	2	15
1	3	14
1	3	16
1	4	13
1	4	12
2	1	27
2	1	25
2	2	27
2	2	26
2	3	25
2	3	24
2	4	23
2	4	23
3	1	31
3	1	32
3	2	27
3	2	24
3	3	25
3	3	26

Table 2
"Fixed" Model
Test of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TX	529.083	2	264.542	**192.394**	**7.65E-10**
Sessions	113.125	3	37.708	**27.424**	**1.17E-05**
TX*Sessions	25.250	6	4.208	3.061	.047
Error	16.500	12	1.375		

Note. The bolded entries differ across the two models.

In a fixed-effects model the MS error (4.208) is used as the denominator in the computation of all *F*s (e.g., 264.542 / 1.375 = 192.394).

Table 3
"Random" Model
Test of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TX	529.083	2	264.542	**62.861**	**9.45E-05**
Error	25.250	6	4.208		
Sessions	113.125	3	37.708	**8.960**	**.012**
Error	25.250	6	4.208		
TX*Sessions	25.250	6	4.208	3.061	.047
Error	16.500	12	1.375		

Note. The bolded entries differ across the two models.

In a random-effects model, the two main effect *F*'s are computed by using the MS of the interaction as the denominator (i.e., 264.542 / 4.208 = 62.861; 37.708 / 4.208 = 8.960; MS interaction = 4.208).

APPENDIX A

SPSS Program to Analyze the "Fixed" and "Random" ANOVA Models

```
SET BLANKS=SYSMIS UNDEFINED=WARN printback=listing .
TITLE 'ANOVA (2001)' .
DATA LIST
FILE= 'a:ANOVA1.txt' FIXED RECORDS=1 TABLE/1
 TX 1 SESSIONS 7 BDI 13-15 .
List variables=all/cases=99999/ format=numbered .

UNIANOVA
 bdi  BY tx sessions
 /RANDOM = tx sessions
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /CRITERIA = ALPHA(.05)
 /DESIGN = tx sessions tx*sessions .
UNIANOVA
 bdi  BY tx sessions
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /CRITERIA = ALPHA(.05)
 /DESIGN = tx sessions tx*sessions .
```

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: INCREASING THE GENERALIZABILITY OF ANOVA RESULTS BY JUDICIOUS SELECTION OF FIXED-, RANDOM-, AND MIXED-EFFECTS MODELS

Author(s): FRANK G. BAUGH

Corporate Source:

Publication Date: 2/1/01

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
FRANK G. BAUGH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X Frank G. Baugh	Position: RES ASSOCIATE
Printed Name: FRANK G. BAUGH	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: 979/845-1335
	Date: 1/24/01

OVER