Confidence Intervals on Generalizability Coefficients for Three-Way Mixed Models and Simulation Study

Hong Zhou[1], Shauna Leonard[2] and Debra Ingram[1]

[1]Department of Mathematics and Statistics, Arkansas State University
P.O.Box 70, State University, AR 72467
hzhou@astate.edu
dingram@astate.edu
[2]Lake Cormorant High School, MS 38641
shauna.leonard@smail.astate.edu

Abstract

Generalizability coefficients (GCs) are commonly used in behavioral measurement and psychometrics when one is interested in the relationship among variables of a common class. The formulas of GCs vary depending on the models. The confidence intervals for GCs and F tests for three-way mixed effects models are obtained in Zhou, Muellerleile, Ingram and Wong (2011). This paper studies the coverage probabilities of proposed confidence intervals using computer simulation. A general algorithm to generate model parameters under constraints is proposed. Simulation study shows that a balanced design with a large n_i, n_j, and n_p for a generalizability coefficient is desirable in order to have a designated confidence interval.

Keywords:*Generalizability coefficients, Intraclass correlation coefficient, ANOVA, random effects model, mixed model*

1 Introduction

Generalizability theory can allow researchers to partition sources of variability that are not captured in classical test theory. It also provides a generalizability coefficient ρ that is analogous to the reliability coefficient in classical test theory. The forms of generalizability coefficients (GCs) are determined by the experimental design or the model the researcher identifies. Thus, the generalizability coefficient is model related, taking different forms under different circumstances (Shrout & Fleiss, 1979).

Previously, Shrout and Fleiss (1979) developed six formulas for one-way models, two-way random effects models, and two-way mixed effects models. McGraw and Wong (1996) developed four additional formulas for the intraclass correlation coefficient and Wong and McGraw (1999) developed the test statistics and confidence intervals for generalizability coefficients based on six different crossed and nested designs for three-way random effects models. However, because in many experiments, not all facets are random, it is useful to look at models with at least one fixed facet (Erilich & Shavelson, 1978). Thus, Zhou, Muellerleile, Ingram and Wong (2011) developed test statistics and confidence intervals for generalizability coefficients for four different three-way mixed effects models in the context of analysis of variance. The context of three-way mixed effects models we discuss here is the two-facet models (Shavelson and Webb, 1991) with one random facet/factor i and

Table 1: Data for Three-Way Designs

		Subject					
		Algebra		Trigonometry		Calculus	
Teacher	Rater	1	2	1	2	1	2
1		4	5	5	5	2	4
2		6	7	7	9	2	6
3		7	7	4	3	2	7
4		6	8	9	9	4	7
5		1	3	5	4	3	1
6		5	4	7	6	5	4
7		5	5	8	8	6	6
8		6	8	9	8	4	6

one fixed facet/factor j. The object of measurement is p which is another random factor. For derivation of mathematical formulas, readers can consult Zhou, Muellerleile, Ingram and Wong (2011).

Thus, one purpose of the current effort is to demonstrate the importance of appropriate model selection by specifying four different models, which yield different results using the same data. In practice, if the wrong model is specified for a particular situation, then the results are likely to be affected and the researchers'conclusions about those data may be imprecise, at best, or entirely incorrect, at worst.

Most formulas are derived from the exact F distribution in Zhou, Muellerleile, Ingram and Wong (2011). However, there are two approximations. Thus, another purpose of the current effort is to propose a Monte Carlo simulation study to evaluate the ability of the proposed confidence intervals for ρ to maintain the designated confidence level.

The remainder of the paper is organized as follows. Section 2 discusses the development of four different mixed effects models and summarizes the estimators and $100(1-\alpha)\%$ confidence intervals for generalizability coefficients. Section 3 states some useful results. Section 4 describes a Monte Carlo simulation study. Section 5 summarizes and discusses the results.

2 Four Three-Way Mixed Effects Models

In a typical reliability study: the evaluation of performance of teachers, n_p teachers are evaluated on n_j subjects of teaching performance by n_i raters. Table 1 displays a hypothetical data set which represents the observations of two raters who evaluate eight teachers' behavior in three academic subjects (algebra, trigonometry, and calculus). In this study, the object of measurement is teacher p which is from a large population and is a random factor. If raters are considered to be exchangeable, then rater i is considered as a random facet. Usually the academic subject j only has a few conditions and is treated as a fixed factor. There are four different kinds of models will be introduced in this section depending on the relationship among three factors: p, i, and j.

2.1 Model 1: $p \times i \times j$ Crossed Design

Each teacher p teaches all three subjects j and is evaluated by two same raters i. Thus, this study is a completely crossed design, denoted by $p\times i\times j$ as model 1. Another application of this model is like an experimental paradigm where researchers may wish to compare the different teaching approaches to examine the extent to which those approaches influence student behavior in some way.

A three-way mixed effects linear model can be use to describe this study as follows:

$$y_{pij} = \mu + \alpha_p + \beta_i + \gamma_j + (\alpha\beta)_{pi} + (\alpha\gamma)_{pj} + (\beta\gamma)_{ij} + e_{pij,e}, \tag{2.1}$$

where y_{pij} is an observation on the pth teacher on jth subject rated by the ith rater; $p = 1, 2, ..., n_p$; $i = 1, 2, ..., n_i$; $j = 1, 2, ..., n_j$. α_p and β_i are random effects of factors p and i, respectively. γ_j is the fixed effect of factor j. $(\alpha\beta)_{pi}$, $(\alpha\gamma)_{pj}$, and $(\beta\gamma)_{ij}$ are the random interaction effects. $e_{pij,e}$ are residuals normally distributed with a mean of 0 and a variance of $\sigma^2_{pij,e}$. Since random effects pij and e are confounding each other, a simple notation σ^2_e is used to denote the variance component $\sigma^2_{pij,e}$. So does for Models 2, 3 and 4. Venn diagrams in Figure 1 show four different models.

2.2 Model 2: $p \times (i : j)$ Partially Nested Design

The number of ratings each rater must complete in Model 1 may be prohibitive in some situations, or the situation may prevent the researcher from using a completely crossed design. In Model 2, there are different raters for each subject. Thus, each teacher p teaches all subjects j and different raters i evaluate different subjects for each teacher. Rater facet i is nested within subject facet j and then crossed with teacher facet p. It is a partially nested design, denoted by $p \times (i : j)$ with two random factors of teacher and rater and a fixed factor of subject. The three-way mixed effects linear model can be represented as follows:

$$y_{pij} = \mu + \alpha_p + \gamma_j + (\beta\gamma)_{i,i:j} + (\alpha\gamma)_{pj} + e_{pi,pij,e}, \tag{2.2}$$

where $p = 1, 2, ..., n_p$; $i = 1, 2, ..., n_i$; $j = 1, 2, ..., n_j$. α_p is random effect of factor p and γ_j is fixed effect of factor j. $(\beta\gamma)_{i,i:j}$ is a random effect of the factor i nested within the factor j confounding with the effect $(i : j)$. $(\alpha\gamma)_{pj}$ is the random interaction effect. $e_{pi,pij,e}$ are residuals normally distributed with a mean of 0 and a variance of σ^2_e.

Compared to a crossed design, some variance components can not be estimated separately in this nested design because of confounding. In Model 2, there must be more raters who are each rating fewer subjects than in Model 1; the result of this difference in design is that variance components associated with the raters i cannot be disaggregated from the teacher p and

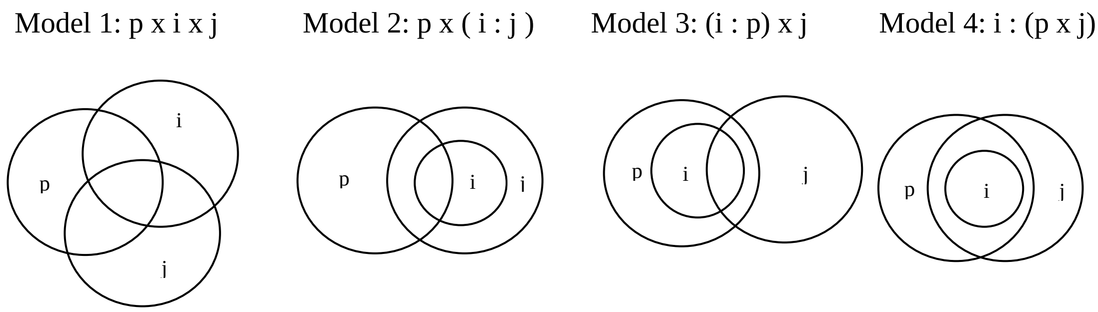

Figure 1: Four Mixed Effects Models for Generalizability Study

subject j, except for the estimate of the variance of the rater within the subject.

Table 2: Observed MS for Three-Way Mixed Effects Models

Observed Mean Squares	Model 1 $p \times i \times j$	Model 2 $p \times (i : j)$	Model 3 $(i : p) \times j$	Model 4 $i : (p \times j)$
MS_p	12.7976	12.7976	12.7976	12.7976
MS_i	6.7500			
MS_j	21.3958	21.3958	21.3958	21.3958
MS_{pi}	1.8452			
MS_{pj}	2.7530	2.7530	2.7530	2.7530
MS_{ij}	3.5625			
MS_e	1.1577	1.3869	1.4583	1.7917
$MS_{i:j}$		4.6250		
$MS_{i:p}$			2.4583	

2.3 Model 3: $(i : p) \times j$ Partially Nested Design

A third example illustrates a different nested design. Each teacher teaches all academic subjects, but different raters evaluate the behavior of different teachers. Thus, rater is nested within teacher. Therefore, it is a partially nested design with the teacher p and rater i as the random factors and subject j as a fixed factor, denoted by $(i : p) \times j$. The three-way mixed effects linear model can be written as follows:

$$y_{pij} = \mu + \alpha_p + \gamma_j + (\alpha\beta)_{i,i:p} + (\alpha\gamma)_{pj} + e_{ij,pij,e}, \tag{2.3}$$

where $p = 1, 2, ..., n_p$; $i = 1, 2, ..., n_i$; $j = 1, 2, ..., n_j$. α_p is a random effect of factor p. γ_j is a fixed effect of factor j. $(\alpha\beta)_{i,i:p}$ is a random effect of the factor i nested within the factor p confounding with the effect i. $(\alpha\gamma)_{pj}$ is the random interaction effects. $e_{ij,pij,e}$ are residuals normally distributed with a mean of 0 and a variance of σ_e^2.

Like Model 2, there are more raters who are each rating a more limited selection of teachers than in Model 1; the result of this difference in design is that variance components associated with the rater cannot be disaggregated from the teacher and subject, except for the estimate of the variance of the rater within the teacher.

2.4 Model 4: $i : (p \times j)$ Partially Nested Design

Our final example illustrates a third possibility for a partially nested design. In this design, each teacher teaches all subjects; however, the rater is different for each combination of the teacher and subject. Thus, the teacher is crossed with the subject. The rater is nested within the combination of teachers and subjects. Therefore, the design is a partially nested design with both teacher p and rater i as random factors and the subject j as a fixed factor, denoted by $i : (p \times j)$. One application of Model 4 might be like end-of-semester student evaluations of their professors.

The three-way mixed effects linear model can be developed as follows:

$$y_{pij} = \mu + \alpha_p + \gamma_j + (\alpha\gamma)_{pj} + e_{i,i:p,i:j,pij,e}, \tag{2.4}$$

where $p = 1, 2, ..., n_p$; $i = 1, 2, ..., n_i$; $j = 1, 2, ..., n_j$. α_p is a random effect of the factor p. γ_j is a fixed effect of factor j. $(\alpha\gamma)_{pj}$ is the random interaction effect. The random effect of the factor i nests within the combination of factors p and j confounding with error e. $e_{i,i:p,i:j,pij,e}$ are residuals normally distributed with a mean of 0 and a variance of σ_e^2.

Like Models 2 and 3, Model 4 requires more raters and precludes isolation of variance components associated with the rater, which is confounded with the teacher and subject. The test statistics and confidence intervals for a

Table 3: 95% Confidence Interval of ρ for Three-Way Mixed Effects Models

Models	Relative			Absolute		
	ρ	Lower	Upper	ρ	Lower	Upper
$p \times i \times j$	0.8558	0.2798	0.9711	0.8167	0.2159	0.9620
$p \times (i:j)$	0.7328	0.4127	0.9304	0.6798	0.3441	0.9123
$(i:p) \times j$	0.8079	0.1301	0.9608	0.8079	0.1301	0.9608
$i:(p \times j)$	0.6719	0.3312	0.9105	0.6719	0.3312	0.9105

particular generalizability coefficient ρ for the three-way mixed effects models (Zhou, Muellerleile, Ingram and Wong, 2011) are summarized in Tables 5, 6, and 7. To help readers appreciate how different mixed model affect the estimation of generalizability coefficient ρ, we apply the various formulas in Tables 5, 6, and 7 to the data in Table 1. Table 2 gives the mean sums of squares and Table 3 shows the estimates of the generalizability coefficients as well as their confidence intervals for four mixed effects models. It is observed that for the same model the estimate of ρ_{abs} is smaller than that of ρ_{rel} because there are more variance components in the denominator of the formula for ρ_{abs}. In addition, ρ is larger for a fully crossed models than that for partially nested models due to the confounding effects.

3 Some Useful Results

An important issue in the analysis of variance is to obtain the expected mean sums of squares (EMS) and corresponding degrees of freedom (*df*) for each effect of interest. The expected mean sums of squares in Table 8 are derived using the rules given in Kirk (1995) and Montgomery (2001). Based on Table 8 the estimators of variance components can be derived. In general, an unbiased estimator of any variance component, say σ_*^2, is a linear combination of mean sums of squares with the form as follows:

$$\hat{\sigma_*^2} = \sum_i k_i (MS_i), \tag{3.1}$$

where k_i is the coefficient of the ith mean sums of squares, MS_i with degrees of freedom df_i. The statistic $\frac{\nu \hat{\sigma_*^2}}{\sigma_*^2}$ has a chi-squared distribution approximately with degrees of freedom of ν, that can be calculated using the following formula:

$$\nu = \frac{(\sum_i k_i (MS_i))^2}{\sum_i k_i^2 (MS_i)^2 / df_i}. \tag{3.2}$$

Then, for the null hypotheses $H_0 : \rho = \rho_0$, an appropriate F-test statistic can be derived. The test statistic is exactly F distributed when there is one mean sum of square in both numerator and denominator. However, the test statistic is approximately F distributed following the procedure proposed by

Satterthwaite (1946) when there are more mean sums of squares in both numerator and denominator. A $100(1-\alpha)\%$ confidence interval for ρ is obtained as well, where ρ can be either relative ρ_{rel} or absolute ρ_{abs}.

As it can been seen that six formulas of confidence intervals have been developed in Tables 5-7, four of them are obtained based on the exact F distribution, and two confidence intervals for ρ_{abs} in Models 1 and 2 are derived from the approximated F distribution. Therefore, a computer simulation study is conducted to examine the coverage probabilities of these two approximated confidence intervals under three-way mixed effects models with a board range of factorial designs.

Another useful result is the generation of the central Chi-square random variable. In the assumption of analysis of variance, a $\frac{df_k * MS_k}{EMS_k}$ term is a independently distributed central Chi-square random variable with df_k degrees of freedom. For example, Model 4: $i:(p\times j)$ in Table 8, the following results are established:

$$\sigma_y^2 = \sigma_p^2 + \sigma_j^2 + \sigma_{pj}^2 + \sigma_e^2 \tag{3.3}$$

$$\chi_1^2 = \frac{(n_p-1)MS_p}{EMS_p} = \frac{(n_p-1)MS_p}{n_i n_j \sigma_p^2 + n_i \sigma_{pj}^2 + \sigma_e^2} \sim \chi^2(n_p-1) \tag{3.4}$$

$$\chi_2^2 = \frac{(n_j-1)MS_j}{EMS_j} = \frac{(n_j-1)MS_j}{n_p n_i \sigma_j^2 + n_i \sigma_{pj}^2 + \sigma_e^2} \sim \chi^2(n_j-1) \tag{3.5}$$

$$\chi_3^2 = \frac{(n_p-1)(n_j-1)MS_{pj}}{EMS_{pj}} = \frac{(n_p-1)(n_j-1)MS_{pj}}{n_i \sigma_{pj}^2 + \sigma_e^2} \sim \chi^2((n_p-1)(n_j-1)) \tag{3.6}$$

$$\chi_4^2 = \frac{n_p n_j (n_i-1)MS_{res}}{EMS_{res}} = \frac{n_p n_j (n_i-1)MS_{res}}{\sigma_e^2} \sim \chi^2(n_p n_j (n_i-1)) \tag{3.7}$$

$\chi_1^2, \chi_2^2, \chi_3^2$, and χ_4^2 follow the central Chi-Square distributions with the degrees of freedom $n_p-1, n_j-1, (n_p-1)(n_j-1), n_p n_j (n_i-1)$, respectively. These results will be used to generate the random variables in the computer simulation study. Similar results can be obtained for other models in Table 8.

4 Computer Simulation Study

Simulation studies are presented to determine the effectiveness of proposed confidence intervals based on the Satterthwaite's method (1946). The results of our simulation studies depend on successful generation of random samples from the designated distribution. According the results in Section 3, we propose a simulation procedure in this section. The advantage of this procedure is to generate central Chi-square random variables and to avoid generating observations y_{pij}, since it is not easy to set the appropriate mean and variance for y_{pij}.

4.1 Calculation of Expected Mean Sums of Squares

Consider Model 4 again. Expected mean squares (EMS) can be calculated using the formulas in Table 8 when the design parameters, n_i, n_j, and n_p, and model parameters σ_p^2, σ_j^2, σ_{pj}^2, and σ_e^2 are chosen.

Without loss of generality, we assume that $\sigma_y^2 = \sigma_p^2 + \sigma_j^2 + \sigma_{pj}^2 + \sigma_e^2 = 1$, and that each variance component can only take equally spaced positive values between 0 and 1, and subject to $L_1 \leq \sigma_p^2 \leq L_2$, $L_1 \leq \sigma_j^2 \leq L_2$, $L_1 \leq \sigma_{pj}^2 \leq L_2$, $L_1 \leq \sigma_e^2 \leq L_2$, where L_1 is the lower bound and L_2 is the upper bound for each variance component.

Let q denote the number of variance components and δ be the increment of value for each variance component in a model. For the model 4, $q = 4$; let $\delta = 0.10$; then, $L_1 = 0.10$ and $L_2 = 0.70$, $m = \frac{L_2 - L_1}{\delta} = 6$; transform the parameters σ_p^2, σ_j^2, σ_{pj}^2, and σ_e^2 into x_1, x_2, x_3, and x_4, respectively. That is,

$$x_1 = \frac{\sigma_p^2 - L_1}{1 - L}, x_2 = \frac{\sigma_j^2 - L_1}{1 - L}, x_3 = \frac{\sigma_{pj}^2 - L_1}{1 - L}, x_1 = \frac{\sigma_e^2 - L_1}{1 - L},$$

where $L = qL_1 = 4 * 0.10 = 0.40$ Therefore, $x_1 + x_2 + x_3 + x_4 = 1$, $x_i = 0, \frac{1}{6}, \frac{2}{6}, \cdots, 1$; $i = 1, 2, 3, 4$.

It is a simplex lattice design (Scheffe, 1958), denoted by $\{q, m\} = \{4, 6\}$. The number of points in a $\{q, m\}$ simplex lattice design is

$$\binom{m+q-1}{m} = \frac{(m+q-1)!}{m!(q-1)!}$$

For model 4, $q = 4$, $m = 6$, therefore, there are 84 distinct sets or combinations of model parameters.

Similarly, if let $\delta = 0.05$ for model 4, then $L_1 = 0.05, L_2 = 0.85$, $m = \frac{0.85-0.05}{0.05} = 16$, thus, there are $\frac{(16+4-1)!}{16!(4-1)!} = 969$ distinct sets of model parameters.

A set of values of n_i, n_j, and n_p gives a particular design. Then, a set of parameters σ_p^2, σ_j^2, σ_{pj}^2, and σ_e^2 will be generated using the simplex lattice design for Model 4. Thus, expected mean squares (EMS) can be calculated using the formulas in Table 8.

4.2 Generation of Random Samples

The mean sums of squares (MS) are needed in order to calculate the confidence interval for a generalizability coefficient for each model. The general procedure to generate random samples for Model 4 is as follows:

1. Define the design parameters n_i, n_j, and n_p. Generate model parameters σ_p^2, σ_j^2, σ_{pj}^2, and σ_e^2 by using the simplex lattice design as described in Section 4.1;

2. Calculate expected mean squares (EMS) using the formulas in Table 8;

3. Generate independent Chi-square random variables $\chi_1^2, \chi_2^2, \chi_3^2$, and χ_4^2 with the degrees of freedom accordingly following the results in Section 3;

4. Calculate mean sums of squares MS_p, MS_j, MS_{pj} and MS_{res}. For Model 4, $MS_p = \chi_1^2 * EMS_p/(n_p - 1)$ and so on;

5. Compute the point estimate and confidence interval of a generalizability coefficient for each sample using the formulas in Tables 5, 6, and 7;

6. Repeat Steps 1-5 for 5,000 times;

7. Calculate the coverage rate by counting the number of times the confidence interval covers the true generalizability coefficient.

5 Results and Conclusions

The simulations were run at all possible combinations of the above parameters. The coverage probability for a given combination of the simulation parameters is estimated by the number of times the confidence bounds include the true ρ divided by the number of simulations, which is 5,000 for each set of parameters. The results are presented in Table 4.

Table 4: Coverage Rates of 95% Confidence Interval of ρ_{abs} for Three-Way Mixed Effects Models

Designs				Model 1			Model 2		
					ρ_{abs}			ρ_{abs}	
n_i	n_j	n_p	ρ_0	Min	Med	Max	Min	Med	Max
2	3	8	0.50	0.9278	0.9520	0.9550	0.9050	0.9377	0.9481
2	3	8	0.60	0.9356	0.9547	0.9587	0.9104	0.9423	0.9483
2	3	8	0.70	0.9434	0.9599	0.9627	0.9148	0.9404	0.9486
2	3	8	0.75	0.9494	0.9620	0.9647	0.9078	0.9421	0.9502
2	3	8	0.80	0.9486	0.9646	0.9668	0.9152	0.9432	0.9504
2	3	8	0.85	0.9544	0.9669	0.9691	0.9130	0.9431	0.9504
12	2	4	0.50	0.9432	0.9500	0.9518	0.9408	0.9478	0.9496
12	2	4	0.60	0.9452	0.9509	0.9542	0.9376	0.9487	0.9509
12	2	4	0.70	0.9454	0.9546	0.9572	0.9410	0.9490	0.9507
12	2	4	0.75	0.9496	0.9574	0.9590	0.9416	0.9486	0.9514
12	2	4	0.80	0.9526	0.9589	0.9612	0.9418	0.9494	0.9516
12	2	4	0.85	0.9536	0.9606	0.9625	0.9390	0.9495	0.9514
5	5	5	0.50	0.9406	0.9499	0.9528	0.9406	0.9486	0.9504
5	5	5	0.60	0.9480	0.9550	0.9570	0.9378	0.9492	0.9516
5	5	5	0.70	0.9530	0.9596	0.9617	0.9416	0.9495	0.9516
5	5	5	0.75	0.9536	0.9633	0.9650	0.9402	0.9499	0.9522
5	5	5	0.80	0.9574	0.9655	0.9672	0.9396	0.9496	0.9519
5	5	5	0.85	0.9612	0.9681	0.9702	0.9416	0.9496	0.9519
10	10	10	0.50	0.9410	0.9481	0.9505	0.9406	0.9497	0.9519
10	10	10	0.60	0.9466	0.9548	0.9569	0.9434	0.9502	0.9519
10	10	10	0.70	0.9526	0.9628	0.9650	0.9410	0.9492	0.9510
10	10	10	0.75	0.9590	0.9664	0.9682	0.9428	0.9498	0.9519
10	10	10	0.80	0.9638	0.9698	0.9714	0.9418	0.9500	0.9524
10	10	10	0.85	0.9670	0.9730	0.9747	0.9430	0.9504	0.9524
30	30	30	0.50	0.9386	0.9491	0.9514	0.9430	0.9496	0.9524
30	30	30	0.60	0.9484	0.9568	0.9592	0.9422	0.9504	0.9522
30	30	30	0.70	0.9584	0.9650	0.9668	0.9396	0.9498	0.9522
30	30	30	0.75	0.9612	0.9686	0.9708	0.9424	0.9498	0.9516

Continued on next page

Table 4 – continued from previous page

Designs				Model 1			Model 2		
					ρ_{abs}			ρ_{abs}	
n_i	n_j	n_p	ρ_0	Min	Med	Max	Min	Med	Max
30	30	30	0.80	0.9653	0.9729	0.9738	0.9394	0.9498	0.9516
30	30	30	0.85	0.9658	0.9764	0.9782	0.9414	0.9501	0.9518
40	40	40	0.50	0.9424	0.9493	0.9508	0.9422	0.9498	0.9520
40	40	40	0.60	0.9524	0.9582	0.9607	0.9434	0.9498	0.9518
40	40	40	0.70	0.9554	0.9648	0.9672	0.9400	0.9494	0.9516
40	40	40	0.75	0.9600	0.9688	0.9710	0.9430	0.9504	0.9522
40	40	40	0.80	0.9628	0.9733	0.9752	0.9422	0.9502	0.9522
40	40	40	0.85	0.9652	0.9765	0.9778	0.9412	0.9496	0.9520
5	25	35	0.50	0.9030	0.9395	0.9551	0.9412	0.9502	0.9522
5	25	35	0.60	0.9294	0.9555	0.9683	0.9414	0.9498	0.9517
5	25	35	0.70	0.9492	0.9718	0.9800	0.9406	0.9498	0.9518
5	25	35	0.75	0.9606	0.9786	0.9850	0.9376	0.9497	0.9516
5	25	35	0.80	0.9674	0.9837	0.9887	0.9404	0.9498	0.9520
5	25	35	0.85	0.9764	0.9891	0.9920	0.9418	0.9500	0.9516
25	5	35	0.50	0.9420	0.9486	0.9516	0.9356	0.9461	0.9482
25	5	35	0.60	0.9506	0.9586	0.9606	0.9384	0.9473	0.9499
25	5	35	0.70	0.9608	0.9684	0.9705	0.9370	0.9471	0.9492
25	5	35	0.75	0.9664	0.9730	0.9747	0.9314	0.9478	0.9502
25	5	35	0.80	0.9688	0.9770	0.9784	0.9396	0.9480	0.9507
25	5	35	0.85	0.9690	0.9812	0.9828	0.9364	0.9483	0.9506
35	25	5	0.50	0.9416	0.9492	0.9516	0.9406	0.9483	0.9504
35	25	5	0.60	0.9442	0.9519	0.9538	0.9404	0.9498	0.9514
35	25	5	0.70	0.9442	0.9534	0.9554	0.9412	0.9494	0.9518
35	25	5	0.75	0.9474	0.9537	0.9556	0.9424	0.9496	0.9514
35	25	5	0.80	0.9476	0.9539	0.9561	0.9436	0.9494	0.9516
35	25	5	0.85	0.9482	0.9560	0.9578	0.9418	0.9496	0.9516
50	50	100	0.50	0.9422	0.9489	0.9508	0.9416	0.9498	0.9520
50	50	100	0.60	0.9546	0.9622	0.9638	0.9424	0.9499	0.9522
50	50	100	0.70	0.9634	0.9736	0.9755	0.9426	0.9501	0.9525
50	50	100	0.75	0.9682	0.9784	0.9802	0.9428	0.9496	0.9522
50	50	100	0.80	0.9694	0.9829	0.9842	0.9434	0.9498	0.9519
50	50	100	0.85	0.9688	0.9863	0.9884	0.9406	0.9500	0.9524
5	10	100	0.50	0.8812	0.9165	0.9401	0.9416	0.9504	0.9522
5	10	100	0.60	0.9152	0.9455	0.9590	0.9402	0.9498	0.9522
5	10	100	0.70	0.9452	0.9659	0.9778	0.9400	0.9504	0.9524
5	10	100	0.75	0.9574	0.9756	0.9829	0.9416	0.9499	0.9519
5	10	100	0.80	0.9696	0.9826	0.9892	0.9416	0.9508	0.9524
5	10	100	0.85	0.9786	0.9889	0.9938	0.9424	0.9503	0.9526

For Model 1 in Table 4, n_i is the levels of factor i; n_j is the levels of factor j; n_p is the levels of factor p. A set of n_i, n_j, and n_p determines a specific design. ρ_0 is a preset value in a null hypothesis, $H_0 : \rho = \rho_0$. Minimum, median, and maximum values of the coverage rates are summarized in Table

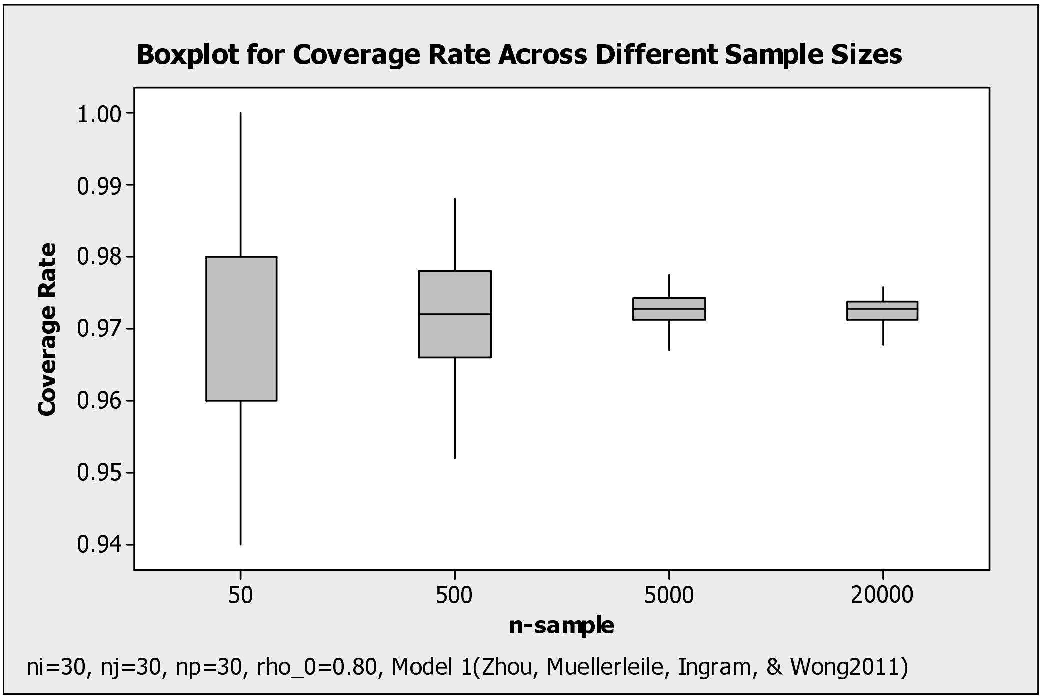

Figure 2: Coverage Rates versus Sample Size

4. n-sample is the sample size. For Model 1, a box-plot in Figure 2, shows that the coverage rate changes as the sample size changes. As the sample size increases, the coverage rate gets closer to the target confidence level 95%. The coverage rate for a sample size of 5,000 is very similar to that of a sample size of 20,000. Thus, the sample size of 5,000 is chosen in the computer simulation for efficiency and economy.

5.1 Coverage Rates across Various Designs

It can be seen that the coverage rate changes as the preset value of ρ_0 as well as across various designs in the Tables 4 for Models 1 and 2. It is observed that the minimum coverage rate increases toward 95% as the ρ_0 increases from 0.50 to 0.85.

When n_i, n_j, and n_p are small, there are a few cases with a minimum coverage rate less than 95%. In addition, there are a few cases that have a minimum coverage rate less than 95% when there are large differences among n_i, n_j, and n_p. The smallest coverage rate in this simulation study is Min = 0.8812, Med = 0.9165,Max = 0.9401 when $n_i = 5, n_j = 10$, $n_p = 100$, and $\rho_0 = 0.50$. Almost all maximum coverage rates are very close to or larger than the designated value of 95% for all designs.

It is very common that one wants to test whether a generalizability coefficient $\rho_0 = 0.80$. For this null hypothesis, the minimum coverage rate is very close to or larger than a nominal value of 95%. Thus, proposed confidence intervals are conservative.

5.2 Conclusions

This article has discussed four three-way mixed effects models in the choice of an appropriate generalizability coefficient ρ in the generalizability and decision study. An example was used to further demonstrate the difference among the estimates of ρ. In practice, it is important to apply the correct mathematical models to a data set in order to reach a valid conclusion in a reliability study. The simulation study shows that a balanced design with a large values of n_i, n_j, and n_p for a generalizability coefficient is desirable in order to have a designated confidence interval.

Acknowledgments

This research is supported by NSF EPSCoR RII Grant (#EPS-0701890,09-EPSCoR-0065) and partially supported by Arkansas State University Faculty Research Fund.

References

[1] Erlich, O., and Shavelson, R. J. (1978). The search for correlations between measures of teacher behavior and student achievement: Measurement problem, conceptualization problem, or both? *Journal of Educational Measurement, 15*, 77-89.

[2] Kirk, R. E. (1995). *Experimental Design: Procedures for the Behavioral Sciences*, Pacific Grove, CA: Brooks/Cole.

[3] McGraw, Kenneth O. and Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.

[4] Montgomery, D.C. (2001) *Design and Analysis of Experiments*,5th Edition (New York: Wiley).

[5] Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics, 2*, 110-114.

[6] Scheffe, H. (1958). Experiments with mistures . *Journal of the Royal Statistical Society. Series B (Methodological), Vol. 20, No.2*, 344-360.

[7] Shavelson, R.J. and Webb, N.W. (1991). *Generalizability Theory: A Primer.* Newbury Park, CA:Sage.

[8] Shrout, P.E. and Fleiss, J. L. (1979). Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

[9] Wong, S.P. and McGraw, Kenneth O. (1999). Confidence intervals and F tests for intraclass correlations based on three-way random effects models. *Educational and Psychological Measurement, 59*, 270-288.

[10] Zhou,H., Muellerleile,P., Ingram, D. and Wong,S.P. (2011). Confidence intervals and F tests for intraclass correlations based on three-way mixed effects models. In press, *Journal of Educational and Behavioral Statistics.*

Appendix

Table 5: Generalizability Coefficients and Their Estimators for Three-Way Mixed Effects Models

Definitions Generalizability Coefficient, ρ	Estimators Generalizability Coefficient, $\hat{\rho}$
Model 1: $p \times i \times j$	
$\rho_{rel} = \frac{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j}}{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j} + \frac{\sigma_{pi}^2}{n_i'} + \frac{\sigma_e^2}{n_j n_i'}}$	$\hat{\rho_{rel}} = \frac{MS_p - MS_{pi}}{MS_p + (m_i - 1)MS_{pi}}$
$\rho_{abs} = \frac{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j}}{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j} + \frac{\sigma_i^2}{n_i'} + \frac{\sigma_{ij}^2}{n_i' n_j} + \frac{\sigma_{pi}^2}{n_i'} + \frac{\sigma_e^2}{n_j n_i'}}$	$\hat{\rho_{abs}} = \frac{MS_p - MS_{pi}}{MS_p + \frac{m_i MS_i}{n_p} - (1 + \frac{m_i}{n_p} - m_i)MS_{pi}}$
Model 2: $p \times (i : j)$	
$\rho_{rel} = \frac{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j}}{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j} + \frac{\sigma_e^2}{n_i'}}$	$\hat{\rho_{rel}} = \frac{MS_p - MS_e}{MS_p + (m_i n_j - 1)MS_e}$
$\rho_{abs} = \frac{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j}}{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j} + \frac{\sigma_{i:j}^2}{n_i'} + \frac{\sigma_e^2}{n_i'}}$	$\hat{\rho_{abs}} = \frac{MS_p - MS_e}{MS_p + \frac{m_i n_j MS_{i:j}}{n_p} - (1 + \frac{m_i n_j}{n_p} - m_i n_j)MS_e}$
Model 3: $(i \times p) \times j$	
$\rho = \frac{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j}}{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j} + \frac{\sigma_{i:p}^2}{n_i'} + \frac{\sigma_e^2}{n_j n_i'}}$	$\hat{\rho} = \frac{MS_p - MS_{i:p}}{MS_p + (m_i - 1)MS_{i:p}}$
Model 4: $i : (p \times j)$	
$\rho = \frac{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j}}{\sigma_p^2 + \frac{\sigma_{pj}^2}{n_j} + \frac{\sigma_e^2}{n_i'}}$	$\hat{\rho} = \frac{MS_p - MS_e}{MS_p + (m_i n_j - 1)MS_e}$

Note: $m_i = n_i/n_i'$, n_i' is the number of raters in the decision study; $\rho_{rel} = \rho_{abs} = \rho$ for Model 3 and 4.

Table 6: Generalizability Coefficients and Their Confidence Intervals for Three-Way Mixed Effects Models 1 and 2

F Test Statistic	Confidence Intervals (l, u)
Model 1: $p \times i \times j$	
$H_0: \rho_{rel} = \rho_0$	$l = \frac{R-F_2}{F_2(m_i-1)+R}$,
$F = \frac{MS_p}{(\frac{m_i\rho_0}{1-\rho_0}+1)MS_{pi}}$	$u = \frac{R-F_1}{F_1(m_i-1)+R}$
where $F(n_p-1,(n_p-1)(n_i-1))$	$F_1 = F_{1-\frac{\alpha}{2}}, F_2 = F_{\frac{\alpha}{2}}$
$H_0: \rho_{abs} = \rho_0$	$l = \frac{n_p(MS_p-F_2MS_{pi})}{n_pMS_P+m_iF_2MS_i+(m_i(n_p-1)-n_p)F_2MS_{pi}}$
$F = \frac{MS_p}{aMS_i+bMS_{pi}}$	$u = \frac{n_p(MS_p-F_1MS_{pi})}{n_pMS_P+m_iF_1MS_i+(m_i(n_p-1)-n_p)F_1MS_{pi}}$
where $\nu = \frac{(aMS_i+bMS_{pi})^2}{\frac{(aMS_i)^2}{n_i-1}+\frac{(bMS_{pi})^2}{(n_p-1)(n_i-1)}}$	$a = \frac{m_i\rho_0}{n_p(1-\rho_0)}, b = 1+\frac{(1-\frac{1}{n_p})m_i\rho_0}{1-\rho_0}$
$F(n_p-1,\nu)$,	$F_1 = F_{1-\frac{\alpha}{2}}, F_2 = F_{\frac{\alpha}{2}}$
Model 2: $p \times (i:j)$	
$H_0: \rho_{rel} = \rho_0$	$l = \frac{R-F_2}{F_2(m_in_j-1)+R}$,
$F = \frac{(1-\rho_0)MS_p}{((m_in_j-1)\rho_0+1)MS_e}$	$u = \frac{R-F_1}{F_1(m_in_j-1)+R}$
where $F(n_p-1, n_j(n_p-1)(n_i-1))$	
$H_0: \rho_{abs} = \rho_0$	$l = \frac{n_p(MS_p-F_2MS_e)}{n_pMS_P+m_in_jF_2MS_{i:j}+(m_in_j(n_p-1)-n_p)F_2MS_e}$
$F = \frac{MS_p}{aMS_{i:j}+bMS_e}$	$u = \frac{n_p(MS_p-F_1MS_e)}{n_pMS_P+m_in_jF_1MS_{i:j}+(m_in_j(n_p-1)-n_p)F_1MS_e}$
where $\nu = \frac{(aMS_{i:j}+bMS_e)^2}{\frac{(aMS_{i:j})^2}{n_j(n_i-1)}+\frac{(bMS_e)^2}{n_j(n_p-1)(n_i-1)}}$	$a = \frac{m_in_j\rho_0}{n_p(1-\rho_0)}, b = 1+\frac{m_in_j\rho_0(1-\frac{1}{n_p})}{1-\rho_0}$
$F(n_p-1,\nu)$	$F_1 = F_{1-\frac{\alpha}{2}}, F_2 = F_{\frac{\alpha}{2}}$

Table 7: Generalizability Coefficients and Their Confidence Intervals for Three-Way Mixed Effects Models 3 and 4

	F Test Statistic	Confidence Intervals (l, u)
Model 3: $(i \times p) \times j$		
	$H_0: \rho = \rho_0$	$l = \frac{R-F_2}{F_2(m_i-1)+R}$,
	$F = \frac{(1-\rho_0)MS_p}{(1+(m_i-1)\rho_0)MS_{i:p}}$	$u = \frac{R-F_1}{F_1(m_i-1)+R}$
	where $F(n_p-1, n_p(n_i-1))$,	$F_1 = F_{1-\frac{\alpha}{2}}, F_2 = F_{\frac{\alpha}{2}}$
Model 4: $i:(p \times j)$		
	$H_0: \rho = \rho_0$	$l = \frac{R-F_2}{F_2(m_in_j-1)+R}$,
	$F = \frac{(1-\rho_0)MS_p}{(1+(m_in_j-1)\rho_0)MS_e}$	$u = \frac{R-F_1}{F_1(m_in_j-1)+R}$
	where $F(n_p-1, n_pn_j(n_i-1))$,	$F_1 = F_{1-\frac{\alpha}{2}}, F_2 = F_{\frac{\alpha}{2}}$

Table 8: Expected MS for Three-Way Random Effects Models

Models	MS	Expected Mean Square (EMS)	Degrees of Freedom (df)
1: $p \times i \times j$			
		Model:$y_{pij} = \mu + \alpha_p + \beta_i + \gamma_j + (\alpha\beta)_{pi} + (\alpha\gamma)_{pj}$ $+(\beta\gamma)_{ij} + e_{res}$	
	MS_p	$EMS_p = n_i n_j \sigma_p^2 + n_j \sigma_{pi}^2 + n_i \sigma_{pj}^2 + \sigma_e^2$	$n_p - 1$
	MS_i	$EMS_i = n_p n_j \sigma_i^2 + n_j \sigma_{pi}^2 + n_p \sigma_{ij}^2 + \sigma_e^2$	$n_i - 1$
	MS_j	$EMS_j = n_p n_i \sigma_j^2 + n_i \sigma_{pj}^2 + n_p \sigma_{ij}^2 + \sigma_e^2$	$n_j - 1$
	MS_{pi}	$EMS_{pi} = n_j \sigma_{pi}^2 + \sigma_e^2$	$(n_p - 1)(n_i - 1)$
	MS_{pj}	$EMS_{pj} = n_i \sigma_{pj}^2 + \sigma_e^2$	$(n_p - 1)(n_j - 1)$
	MS_{ij}	$EMS_{ij} = n_p \sigma_{ij}^2 + \sigma_e^2$	$(n_i - 1)(n_j - 1)$
	MS_{res}	$EMS_{res} = \sigma_e^2$	$(n_i - 1)(n_j - 1)(n_p - 1)$
2: $p \times (i : j)$			
		Model:$y_{pij} = \mu + \alpha_p + \gamma_j + (\beta\gamma)_{i:j} + (\alpha\gamma)_{pj} + e_{res}$	
	MS_p	$EMS_p = n_i n_j \sigma_p^2 + n_i \sigma_{pj}^2 + \sigma_e^2$	$n_p - 1$
	MS_j	$EMS_j = n_p n_i \sigma_j^2 + n_p \sigma_{i:j}^2 + \sigma_e^2$	$n_j - 1$
	$MS_{i:j}$	$EMS_{i:j} = n_p \sigma_{i:j}^2 + \sigma_e^2$	$(n_i - 1)n_j$
	MS_{pj}	$EMS_{pj} = n_i \sigma_{pj}^2 + \sigma_e^2$	$(n_p - 1)(n_j - 1)$
	MS_{res}	$EMS_{res} = \sigma_e^2$	$(n_i - 1)n_j(n_p - 1)$
3: $(i : p) \times j$			
		Model:$y_{pij} = \mu + \alpha_p + (\alpha\beta)_{i:p} + \gamma_j + (\alpha\gamma)_{pj} + e_{res}$	
	MS_p	$EMS_p = n_i n_j \sigma_p^2 + n_i \sigma_{pj}^2 + n_j \sigma_{i:p}^2 + \sigma_e^2$	$n_p - 1$
	MS_j	$EMS_j = n_p n_i \sigma_j^2 + n_i \sigma_{pj}^2 + \sigma_e^2$	$n_j - 1$
	$MS_{i:p}$	$EMS_{i:p} = n_j \sigma_{i:p}^2 + \sigma_e^2$	$(n_i - 1)n_p$
	MS_{pj}	$EMS_{pj} = n_i \sigma_{pj}^2 + \sigma_e^2$	$(n_p - 1)(n_j - 1)$
	MS_{res}	$EMS_{res} = \sigma_e^2$	$(n_i - 1)n_p(n_j - 1)$
4: $i : (p \times j)$			
		Model: $y_{pij} = \mu + \alpha_p + \gamma_j + (\alpha\gamma)_{pj} + e_{res}$	
	MS_p	$EMS_p = n_i n_j \sigma_p^2 + n_i \sigma_{pj}^2 + \sigma_e^2$	$n_p - 1$
	MS_j	$EMS_j = n_p n_i \sigma_j^2 + n_i \sigma_{pj}^2 + \sigma_e^2$	$n_j - 1$
	MS_{pj}	$EMS_{pj}=n_i \sigma_{pj}^2+\sigma_e^2$	$(n_p - 1)(n_j - 1)$
	MS_{res}	$EMS_{res} = \sigma_e^2$	$(n_i - 1)n_p(n_j - 1)$