Hierarchical Modeling for Bayesian Approach in Generalizability Theory

一般化可能性理論におけるベイズ的方法のための階層的モデル構成

Yasuharu OKAMOTO
岡　本　安　晴
（日本女子大学人間社会学部心理学科）

［**Abstract**］ A natural correspondence between generalizability theory and Bayesian data analysis has been observed. To conduct Bayesian data analyses, explicit stochastic models of generalizability theory were presented with hierarchical priors, which represent random effects, and algorithms of Markov chain Monte Carlo (MCMC) for Bayesian analysis were proposed. To obtain stable posterior distributions of variances, inverse-gamma distributions of variances were employed. The proposed algorithms treat one-facet and two-facet designs whose facets are assumed to be random effects. Successful applications of the proposed Bayesian methods to hypothetical data sets indicate the usefulness and importance of the proposed Bayesian approach in generalizability theory.

Key words: generalizability, reliability, Bayesian analysis, MCMC, hierarchical model

Introduction

A natural correspondence between generalizability theory and Bayesian approach can be pointed out, and this paper presents examples of simple modeling of random facets (factors) in generalizability theory, based on which Bayesian approaches were successfully conducted using hierarchical modeling. Generalizability theory (G theory) consists of two studies, generalizability study (G study) and decision study (D study). Bayesian data analysis consists of three steps; setting up a full probability model, conditioning on observed data, and evaluation (Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin, 2014). G study corresponds to setting up a full probability model and conditioning on observed data, and D study to evaluation. Hence, the Bayesian approach is natural to generalizability theory.

Usually, G theory stands on variance decomposition, which is derived from an ANOVA model (Brennan, 2001, 2011; Cardinet, Johnson, and Pini, 2010; Glas, 2012; Kreiter, 2010; LoPilato, Carter, and Wang, 2014; Shavelson and Webb, 1991). For example, in cases of one-facet design, variance of X_{ti}, a score of the t-th target on the i-th item, is decomposed as follows:

$$\sigma_X^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2, \tag{1}$$

where σ_X^2 is the overall variance of scores, σ_α^2, variance due to targets, σ_β^2, variance due to test items, and σ_e^2, variance due to residuals. The decomposition corresponds to that of the sum of squares (SS) in ANOVA and expectations of SSs are derived (Kirk, 1995; Myers and Well, 2003; Winer, Brown, and Michels, 1991). Based on these expectations, point estimates of variances are calculated. But, estimated variances by expectations of SSs can be negative (Cardinet, Johnson, and Pini, 2010).

On the other hand, Bayesian approaches need formulae of probability distributions of variances, other than formulae of point estimations of variances by expectations, and can keep variances within nonnegative values by prior distributions. In cases of random facets (factors), probability distributions of variances can be successfully treated by hierarchical models (Kruschke, 2011). In the following sections, models of one-facet and two-facet designs, whose facets are represented as random factors in ANOVA models, are presented and successful applications of Bayesian approach by MCMC are reported. The MCMC uses Metropolis-within-Gibbs (Robert and Casella, 2010a, 2010b) or the component-wise version of the Metropolis–Hastings algorithm (Gamerman and Lopes, 2006), which makes adaptive stage of MCMC simpler and calculations of samples from posterior distributions easier.

Models and Algorithms (G study)

This study treats one-facet and two-facet designs, where facets are random factors. As the simplest model, first consider a one-facet design.

One-facet design

According to ANOVA, we have

$$X_{ti} = \mu_G + \alpha_t + \beta_i + e_{ti}, \tag{2}$$

where μ_G is the grand mean, α_t and β_i are effects of target t and item i, respectively, and e_{ti} is residual. α_t, β_i, and e_{ti} are assumed to be random and independent of each other and to have normal distributions:

$$\alpha_t \sim N\left(0, \sigma_\alpha^2\right),$$

$$\beta_i \sim N\left(0, \sigma_\beta^2\right),$$

and

$$e_{ti} \sim N\left(0, \sigma_e^2\right),$$

Set

$$\bar{X}_{t\bullet} = \frac{1}{n_i}\sum_{i=1}^{n_i} X_{ti} = \mu_G + \alpha_t + \frac{1}{n_i}\sum_{i=1}^{n_i}\beta_i + \frac{1}{n_i}\sum_{i=1}^{n_i} e_{ti},$$

where n_i is the number of items.

Under the independence condition (Hogg, McKean, and Craig, 2005), we have

$$V\left(\bar{X}_{t\bullet}\right)=V\left(\alpha_t\right)+\frac{1}{n_i^2}\sum_{i=1}^{n_i}V\left(\beta_i\right)+\frac{1}{n_i^2}\sum_{i=1}^{n_i}V\left(e_{ti}\right)$$
$$=\sigma_\alpha^2+\frac{\sigma_\beta^2}{n_i}+\frac{\sigma_e^2}{n_i}, \quad (3)$$

where $V(Y)$ represents variance of Y. Compare (3) with (1). Although (1) shows the decomposition of variance simply as the sum of component variances, which corresponds to the decomposition of the sum of squares in ANOVA, (3) shows the decomposition of variance explicitly as linear combination of variances of random effects, each weighted by the inverse of the number of levels of each facet (factor). Variances σ_α^2 , σ_β^2 , and σ_e^2 can be estimated by a Bayesian method with Markov chain Monte Carlo (MCMC).

A Bayesian method presupposes a stochastic model. From (2), we set

$$P\left(X_{ti}\middle|\mu_G,\alpha_t,\beta_i,\sigma_\alpha^2,\sigma_\beta^2,\sigma_e^2,\right)=\phi\left(\left\{X_{ti}-\left(\mu_G+\alpha_t+\beta_i\right)\right\}/\sigma_e\right),$$

where $\phi_0(z)$ is the probability density function of the standard normal distribution.

Put

$$\boldsymbol{\alpha}=\left(\alpha_1,\cdots,\alpha_{n_t}\right),\ \boldsymbol{\beta}=\left(\beta_1,\cdots,\beta_{n_i}\right),\ \boldsymbol{X}=\left(X_{11},\cdots,X_{n_t n_i}\right),$$

and

$$n_t = \text{ the number of targets,}$$

then we have the posterior distribution

$$P\left(\mu_G,\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\alpha^2,\sigma_\beta^2,\sigma_e^2\middle|\boldsymbol{X}\right)\propto P\left(\mu_G,\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\alpha^2,\sigma_\beta^2,\sigma_e^2\right)P\left(\boldsymbol{X}\middle|\mu_G,\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\alpha^2,\sigma_\beta^2,\sigma_e^2\right), \quad (4)$$

where

$$P\left(\boldsymbol{X}\middle|\mu_G,\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\alpha^2,\sigma_\beta^2,\sigma_e^2\right)=\prod_{t=1}^{n_t}\prod_{i=1}^{n_i}P\left(X_{ti}\middle|\mu_G,\alpha_t,\beta_i,\sigma_\alpha^2,\sigma_\beta^2,\sigma_e^2\right).$$

Set the following hierarchical prior (Kruschke, 2011)

$$P\left(\mu_G,\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\alpha^2,\sigma_\beta^2,\sigma_e^2\right)=P(\mu_G)\prod_{t=1}^{n_t}P(\alpha_t|\sigma_\alpha^2)\prod_{i=1}^{n_i}P\left(\beta_i\middle|\sigma_\beta^2\right)P(\sigma_\alpha^2)P\left(\sigma_\beta^2\right)P(\sigma_e^2),$$

where

$$\mu_G\sim U(-C,C),$$

$$\alpha_t\sim N(0,\sigma_\alpha^2), \quad (5)$$

$$\beta_i\sim N\left(0,\sigma_\beta^2\right), \quad (6)$$

$$\sigma_\alpha^2 \sim Inv - Gamma(a_\alpha, b_\alpha),$$

$$\sigma_\beta^2 \sim Inv - Gamma\left(a_\beta, b_\beta\right),$$

$$\sigma_e^2 \sim Inv - Gamma(a_e, b_e).$$

$\mathrm{U}(-\mathrm{C}, \mathrm{C})$ denotes a uniform distribution in $(-\mathrm{C}, \mathrm{C})$; C is set as sufficiently great so that a generated sample does not jump out of the range. In many cases, vague priors are reasonable (Lunn, Jackson, Best, Thomas, and Spiegelhalter, 2013, p. 82). Alcalá-Quintana and García-Pérez (2004) recommended a uniform distribution as a prior distribution of a position parameter. Kingdom and Prins (2010) used uniform priors limited to bounded regions to calculate the posterior distributions of the position and scale parameters of PFs.

(5) and (6) represent α_t and β_i as random effects. Prior distributions of σ_α^2 and σ_β^2 are set as inverse gamma distributions. Although a bounded uniform distribution is recommended as a noninformative prior for a variance parameter σ (Carlin and Louis, 2009; Gelman, 2006), in some cases, uniform prior distributions of variances make posterior distributions unstable (Okamoto, 2013). A gamma distribution is used as a conjugate prior distribution for the inverse of variance (Kruschke, 2011), and in this case, the distribution of variance is an inverse-gamma distribution. Parameters of inverse-gamma distributions are set to weakly reflect prior information.

Posterior distribution (4) can be estimated by the following Metropolis-within-Gibbs (Robert and Casella, 2010a, 2010b) or the component-wise version of the Metropolis–Hastings algorithm (Gamerman and Lopes, 2006). The algorithm uses normal proposal distributions. A proposed value for a parameter whose value is restricted to positive is constrained to positive by the prior distribution. Cycles of MCMC steps proceed as follows:

Step 0. Set initial values $\mu_G^{(0)}$, $\alpha_t^{(0)}$, $\beta_i^{(0)}$, $\sigma_\alpha^{2(0)}$, $\sigma_\beta^{2(0)}$, $\sigma_e^{2(0)}$, where $\mu_G^{(s)}$ and so on denote values at the s-th iteration.

Set $s \leftarrow 0$.

Step 1. Draw a sample $y \sim N\left(\mu_G^{(s)}, \sigma_{\mu_G}^2\right)$

Calculate acceptance probability

$$a = \min\left\{1, \frac{P\left(y, \alpha_1^{(s)}, \cdots, \sigma_e^{2(s)} \middle| X\right)\phi\left(\mu_G^{(s)}; y, \sigma_{\mu_G}^2\right)}{P\left(\mu_G^{(s)}, \alpha_1^{(s)}, \cdots, \sigma_e^{2(s)} \middle| X\right)\phi\left(y; \mu_G^{(s)}, \sigma_{\mu_G}^2\right)}\right\},$$

where $\phi(x; \mu, \sigma^2)$ is the probability density function of a normal distribution with mean μ and variance σ^2, i.e.,

$$\phi(x; \mu, \sigma^2) = \phi_0((x - \mu)/\sigma).$$

Set $\mu_G^{(s+1)} \leftarrow y$ with probability a; otherwise set $\mu_G^{(s+1)} \leftarrow \mu_G^{(s)}$.

Step 2. Repeat steps 2a to 2c for $t = 1$ to n_t.

2a. Draw a sample $y \sim N\left(\alpha_t^{(s)}, \sigma_{g\alpha}^2\right)$.

2b. Calculate acceptance probability a

$$a = \min\left\{1, \frac{P\left(\cdots,\alpha_{t-1}^{(s+1)},y,\alpha_{t+1}^{(s)}\cdots\middle|\boldsymbol{X}\right)\phi\left(\alpha_t^{(s)};y,,\sigma_{g\alpha}^2\right)}{P\left(\cdots,\alpha_{t-1}^{(s+1)},\alpha_t^{(s)},\alpha_{t+1}^{(s)}\cdots\middle|\boldsymbol{X}\right)\phi\left(y;\alpha_t^{(s)},\sigma_{g\alpha}^2\right)}\right\}$$

2c. Set $\alpha_t^{(s+1)} \leftarrow y$ with probability a; otherwise set $\alpha_t^{(s+1)} \leftarrow \alpha_t^{(s)}$.

Step 3. Repeat steps 3a to 3c for $i = 1$ to n_i.

3a. Draw a sample $y \sim N\left(\beta_i^{(s)}, \sigma_{g\beta}^2\right)$.

3b. Calculate acceptance probability a

$$a = \min\left\{1, \frac{P\left(\cdots,\beta_{i-1}^{(s+1)},y,\beta_{i+1}^{(s)}\cdots\middle|\boldsymbol{X}\right)\phi\left(\beta_i^{(s)};y,,\sigma_{gi}^2\right)}{P\left(\cdots,\beta_{i-1}^{(s+1)},\beta_i^{(s)},\beta_{i+1}^{(s)}\cdots\middle|\boldsymbol{X}\right)\phi\left(y;\beta_i^{(s)},\sigma_{g\beta}^2\right)}\right\}$$

3c. Set $\beta_i^{(s+1)} \leftarrow y$ with probability a; otherwise set $\beta_i^{(s+1)} \leftarrow \beta_i^{(s)}$.

Step 4. Draw a sample $y \sim N\left(\sigma_\alpha^{2(s)}, \sigma_{\sigma_\alpha}^2\right)$

Calculate acceptance probability a

$$a = \min\left\{1, \frac{P\left(\cdots,\beta_{n_i}^{(s+1)},y,\sigma_\beta^{2(s)},\cdots\middle|\boldsymbol{X}\right)\phi\left(\sigma_\alpha^{2(s)};y,\sigma_{\sigma_\alpha}^2\right)}{P\left(\cdots,\beta_{n_i}^{(s+1)},\sigma_\alpha^{2(s)},\sigma_\beta^{2(s)},\cdots\middle|\boldsymbol{X}\right)\phi\left(y;\sigma_\alpha^{2(s)},\sigma_{\sigma_\alpha}^2\right)}\right\}$$

Set $\sigma_\alpha^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_\alpha^{2(s+1)} \leftarrow \sigma_\alpha^{2(s)}$.

Step 5. Draw a sample $y \sim N\left(\sigma_\beta^{2(s)}, \sigma_{\sigma_\beta}^2\right)$

Calculate acceptance probability a

$$a = \min\left\{1, \frac{P\left(\cdots,\sigma_\alpha^{2(s+1)},y,\sigma_e^{2(s)}\middle|\boldsymbol{X}\right)\phi\left(\sigma_\beta^{2(s)};y,\sigma_{\sigma_\beta}^2\right)}{P\left(\cdots,\sigma_\alpha^{2(s+1)},\sigma_\beta^{2(s)},\sigma_e^{2(s)}\middle|\boldsymbol{X}\right)\phi\left(y;\sigma_\beta^{2(s)},\sigma_{\sigma_\beta}^2\right)}\right\}$$

Set $\sigma_\beta^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_\beta^{2(s+1)} \leftarrow \sigma_\beta^{2(s)}$.

Step 6. Draw a sample $y \sim N\left(\sigma_e^{2(s)}, \sigma_{\sigma_e}^2\right)$

Calculate acceptance probability a

$$a = \min\left\{1, \frac{P\left(\cdots, \sigma_\beta^{2(s+1)}, y \middle| X\right)\phi\left(\sigma_e^{2(s)}; y, \sigma_{\sigma_e}^2\right)}{P\left(\cdots, \sigma_\beta^{2(s+1)}, \sigma_e^{2(s)} \middle| X\right)\phi\left(y; \sigma_e^{2(s)}, \sigma_{\sigma_e}^2\right)}\right\}$$

Set $\sigma_e^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_e^{2(s+1)} \leftarrow \sigma_e^{2(s)}$.

Step 7. Set $s \leftarrow s + 1$.

If s does not reach the number set as the total number of iterations in MCMC, return to step 1.

Using samples $\mu_G^{(s)}$, $\alpha_t^{(s)}$, $\beta_i^{(s)}$, $\sigma_\alpha^{2(s)}$, $\sigma_\beta^{2(s)}$, and $\sigma_e^{2(s)}$, we can estimate various posterior distributions, on which D study will be conducted.

Two-facet design

Let X_{tio} be a score of target t on item i by observer o ; then according to ANOVA, we have

$$X_{tio} = \mu_G + \alpha_t + \beta_i + \gamma_o + (\alpha\beta)_{ti} + (\alpha\gamma)_{to} + (\beta\gamma)_{io} + e_{tio}, \tag{7}$$

where μ_G is the grand mean, and α_t, β_i, and γ_o are main effects of target t, item i, and observer o. $(\alpha\beta)_{ti}$, $(\alpha\gamma)_{to}$, and $(\beta\gamma)_{io}$ are interaction effects of target t by item i, target t by observer o, and item i by observer o, respectively, and e_{tio} is residual. All effects are assumed to be independent of each other, and to have normal distributions:

$$\alpha_t \sim N(0, \sigma_\alpha^2), \qquad \beta_i \sim N\left(0, \sigma_\beta^2\right), \qquad \gamma_o \sim N\left(0, \sigma_\gamma^2\right),$$

$$(\alpha\beta)_{ti} \sim N\left(0, \sigma_{\alpha\beta}^2\right), \qquad (\alpha\gamma)_{to} \sim N\left(0, \sigma_{\alpha\gamma}^2\right), \qquad (\beta\gamma)_{io} \sim N\left(0, \sigma_{\beta\gamma}^2\right),$$

$$e_{tio} \sim N(0, \sigma_e^2).$$

Set

$$\begin{aligned}\bar{X}_{t\cdot\cdot} &= \frac{1}{n_i n_o}\sum_{i=1}^{n_i}\sum_{o=1}^{n_o} X_{tio} \\ &= \mu_G + \alpha_t + \frac{1}{n_i}\sum_{i=1}^{n_i}\beta_i + \frac{1}{n_o}\sum_{o=1}^{n_o}\gamma_o \\ &\quad + \frac{1}{n_i}\sum_{i=1}^{n_i}(\alpha\beta)_{ti} + \frac{1}{n_o}\sum_{o=1}^{n_o}(\alpha\gamma)_{to} + \frac{1}{n_i n_o}\sum_{i=1}^{n_i}\sum_{o=1}^{n_o}(\beta\gamma)_{io}\end{aligned}$$

$$+\frac{1}{n_i n_o}\sum_{i=1}^{n_i}\sum_{o=1}^{n_o} e_{tio},$$

where n_o is the number of observers. Under the independence condition, we have

$$V(\bar{X}_{t\cdot\cdot}) = V(\alpha_t) + \frac{1}{n_i^2}\sum_{i=1}^{n_i} V(\beta_i) + \frac{1}{n_o^2}\sum_{o=1}^{n_o} V(\gamma_o) + \frac{1}{n_i^2}\sum_{i=1}^{n_i} V((\alpha\beta)_{ti})$$

$$+\frac{1}{n_o^2}\sum_{o=1}^{n_o} V((\alpha\gamma)_{to}) + \frac{1}{n_i^2 n_o^2}\sum_{i=1}^{n_i}\sum_{o=1}^{n_o} V((\beta\gamma)_{io})$$

$$+\frac{1}{n_i^2 n_o^2}\sum_{i=1}^{n_i}\sum_{o=1}^{n_o} V(e_{tio})$$

$$= \sigma_\alpha^2 + \frac{\sigma_\beta^2}{n_i} + \frac{\sigma_\gamma^2}{n_o} + \frac{\sigma_{\alpha\beta}^2}{n_i} + \frac{\sigma_{\alpha\gamma}^2}{n_o} + \frac{\sigma_{\beta\gamma}^2}{n_i n_o} + \frac{\sigma_e^2}{n_i n_o} \quad (8)$$

Compare (8) with (1). Decomposition of the variance in (8) is explicitly represented as a linear combination of variances of random effects, each weighted by the inverse of the number of levels of each facet. The variances can be estimated by a Bayesian method with MCMC.

From (7), we have

$$P\left(X_{tio} \middle| \mu_G, \alpha_t, \beta_i, \gamma_o, (\alpha\beta)_{ti}, (\alpha\gamma)_{to}, (\beta\gamma)_{io}, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_{\alpha\beta}^2, \sigma_{\alpha\gamma}^2, \sigma_{\beta\gamma}^2, \sigma_e^2\right)$$

$$= \phi(\{X_{tio} - (\mu_G + \alpha_t + \beta_i + \gamma_o + (\alpha\beta)_{ti} + (\alpha\gamma)_{to} + (\beta\gamma)_{io})\}/\sigma_e).$$

Put

$$\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{n_t}),\ \boldsymbol{\beta} = (\beta_1, \cdots, \beta_{n_i}),\ \boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_{n_o}),$$

$$(\boldsymbol{\alpha\beta}) = ((\alpha\beta)_{11}, \cdots, (\alpha\beta)_{n_t n_i}),\ (\boldsymbol{\alpha\gamma}) = ((\alpha\gamma)_{11}, \cdots, (\alpha\gamma)_{n_t n_0}),$$

$$(\boldsymbol{\beta\gamma}) = ((\beta\gamma)_{11}, \cdots, (\beta\gamma)_{n_i n_o}), \text{ and } \boldsymbol{X} = (X_{111}, \cdots, X_{n_t n_i n_o}),$$

then we have the posterior distribution

$$P\left(\mu_G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, (\boldsymbol{\alpha\beta}), (\boldsymbol{\alpha\gamma}), (\boldsymbol{\beta\gamma}), \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_{\alpha\beta}^2, \sigma_{\alpha\gamma}^2, \sigma_{\beta\gamma}^2, \sigma_e^2 \middle| \boldsymbol{X}\right)$$

$$\propto P\left(\mu_G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, (\boldsymbol{\alpha\beta}), (\boldsymbol{\alpha\gamma}), (\boldsymbol{\beta\gamma}), \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_{\alpha\beta}^2, \sigma_{\alpha\gamma}^2, \sigma_{\beta\gamma}^2, \sigma_e^2\right)$$

$$\times P\left(\boldsymbol{X} \middle| \mu_G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, (\boldsymbol{\alpha\beta}), (\boldsymbol{\alpha\gamma}), (\boldsymbol{\beta\gamma}), \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_{\alpha\beta}^2, \sigma_{\alpha\gamma}^2, \sigma_{\beta\gamma}^2, \sigma_e^2\right), \quad (9)$$

where

$$P\left(\boldsymbol{X} \middle| \mu_G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, (\boldsymbol{\alpha\beta}), (\boldsymbol{\alpha\gamma}), (\boldsymbol{\beta\gamma}), \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_{\alpha\beta}^2, \sigma_{\alpha\gamma}^2, \sigma_{\beta\gamma}^2, \sigma_e^2\right)$$

$$= \prod_{t=1}^{n_t} \prod_{i=1}^{n_i} \prod_{o=1}^{n_o} P\left(X_{tio} \middle| \mu_G, \alpha_t, \beta_i, \gamma_o, (\alpha\beta)_{ti}, (\alpha\gamma)_{to}, (\beta\gamma)_{io}, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_{\alpha\beta}^2, \sigma_{\alpha\gamma}^2, \sigma_{\beta\gamma}^2, \sigma_e^2\right).$$

Set the following hierarchical prior (Kruschke, 2011)

$$P\left(\mu_G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, (\boldsymbol{\alpha\beta}), (\boldsymbol{\alpha\gamma}), (\boldsymbol{\beta\gamma}), \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_{\alpha\beta}^2, \sigma_{\alpha\gamma}^2, \sigma_{\beta\gamma}^2, \sigma_e^2\right)$$

$$= P(\mu_G) \prod_{t=1}^{n_t} P(\alpha_t | \sigma_\alpha^2) \prod_{i=1}^{n_i} P\left(\beta_i \middle| \sigma_\beta^2\right) \prod_{o=1}^{n_o} P\left(\gamma_o \middle| \sigma_\gamma^2\right)$$

$$\times \prod_{t=1}^{n_t} \prod_{i=1}^{n_i} P\left((\alpha\beta)_{ti} \middle| \sigma_{\alpha\beta}^2\right) \prod_{t=1}^{n_t} \prod_{o=1}^{n_o} P\left((\alpha\gamma)_{to} \middle| \sigma_{\alpha\gamma}^2\right) \prod_{i=1}^{n_i} \prod_{o=1}^{n_o} P\left((\beta\gamma)_{io} \middle| \sigma_{\beta\gamma}^2\right)$$

$$\times P(\sigma_\alpha^2) P\left(\sigma_\beta^2\right) P\left(\sigma_\gamma^2\right) P\left(\sigma_{\alpha\beta}^2\right) P\left(\sigma_{\alpha\gamma}^2\right) P\left(\sigma_{\beta\gamma}^2\right) P(\sigma_e^2),$$

where

$$\mu_G \sim U(-C, C), \qquad \alpha_t \sim N(0, \sigma_\alpha^2), \qquad \beta_i \sim N\left(0, \sigma_\beta^2\right), \qquad \gamma_o \sim N\left(0, \sigma_\gamma^2\right),$$

$$(\alpha\beta)_{ti} \sim N\left(0, \sigma_{\alpha\beta}^2\right), \qquad (\alpha\gamma)_{to} \sim N\left(0, \sigma_{\alpha\gamma}^2\right), \qquad (\beta\gamma)_{io} \sim N\left(0, \sigma_{\beta\gamma}^2\right),$$

$$\sigma_\alpha^2 \sim Inv - Gamma(a_\alpha, b_\alpha), \qquad \sigma_\beta^2 \sim Inv - Gamma\left(a_\beta, b_\beta\right),$$

$$\sigma_\gamma^2 \sim Inv - Gamma\left(a_\gamma, b_\gamma\right), \qquad \sigma_{\alpha\beta}^2 \sim Inv - Gamma\left(a_{\alpha\beta}, b_{\alpha\beta}\right),$$

$$\sigma_{\alpha\gamma}^2 \sim Inv - Gamma\left(a_{\alpha\gamma}, b_{\alpha\gamma}\right), \qquad \sigma_{\beta\gamma}^2 \sim Inv - Gamma\left(a_{\beta\gamma}, b_{\beta\gamma}\right),$$

$$\sigma_e^2 \sim Inv - Gamma(a_e, b_e).$$

These hierarchical priors are selected for the same reason as in one-facet design.

Posterior distribution (9) can be estimated by essentially the same algorithm as one-facet design. In the following, the algorithm is presented in a concise way. Acceptance probabilities are not presented explicitly, because their expressions can be inferred easily by comparison with those of one-facet design.

Step 0. Set initial values $\mu_G^{(0)}$, $\alpha_t^{(0)}$, $\beta_i^{(0)}$, $\gamma_o^{(0)}$, $(\alpha\beta)_{ti}^{(0)}$, $(\alpha\gamma)_{to}^{(0)}$, $(\beta\gamma)_{io}^{(0)}$, $\sigma_\alpha^{2(0)}$, $\sigma_\beta^{2(0)}$, $\sigma_\gamma^{2(0)}$, $\sigma_{\alpha\beta}^{2(0)}$, $\sigma_{\alpha\gamma}^{2(0)}$, $\sigma_{\beta\gamma}^{2(0)}$, $\sigma_e^{2(0)}$, where $\mu_G^{(s)}$ and so on denote values at the s-th iteration.

Set $s \leftarrow 0$.

Step 1. Draw a sample $y \sim N\left(\mu_G^{(s)}, \sigma_{\mu_G}^2\right)$.

Calculate acceptance probability a.

Set $\mu_G^{(s+1)} \leftarrow y$ with probability a; otherwise set $\mu_G^{(s+1)} \leftarrow \mu_G^{(s)}$.

Step 2. Repeat steps 2a to 2c for $t = 1$ to n_t.

2a. Draw a sample $y \sim N\left(\alpha_t^{(s)}, \sigma_{g\alpha}^2\right)$.

2b. Calculate acceptance probability a.

2c. Set $\alpha_t^{(s+1)} \leftarrow y$ with probability a; otherwise set $\alpha_t^{(s+1)} \leftarrow \alpha_t^{(s)}$.

Step 3. Repeat steps 3a to 3c for $i = 1$ to n_i.

3a. Draw a sample $y \sim N\left(\beta_i^{(s)}, \sigma_{g\beta}^2\right)$.

3b. Calculate acceptance probability a.

3c. Set $\beta_i^{(s+1)} \leftarrow y$ with probability a; otherwise set $\beta_i^{(s+1)} \leftarrow \beta_i^{(s)}$.

Step 4. Repeat steps 4a to 4c for $\mathrm{o} = 1$ to n_o.

4a. Draw a sample $y \sim N\left(\gamma_o^{(s)}, \sigma_{g\gamma}^2\right)$.

4b. Calculate acceptance probability a.

4c. Set $\gamma_o^{(s+1)} \leftarrow y$ with probability a; otherwise set $\gamma_o^{(s+1)} \leftarrow \gamma_o^{(s)}$.

Step 5. Repeat steps 5a to 5c for $t = 1$ to n_t.

Repeat steps 5a to 5c for $i = 1$ to n_i.

5a. Draw a sample $y \sim N\left((\alpha\beta)_{ti}^{(s)}, \sigma_{g\alpha\beta}^2\right)$.

5b. Calculate acceptance probability a.

5c. Set $(\alpha\beta)_{ti}^{(s+1)} \leftarrow y$ with probability a;

otherwise set $(\alpha\beta)_{ti}^{(s+1)} \leftarrow (\alpha\beta)_{ti}^{(s)}$.

Step 6. Repeat steps 6a to 5c for $t = 1$ to n_t.

Repeat steps 6a to 5c for $o = 1$ to n_o.

6a. Draw a sample $y \sim N\left((\alpha\gamma)_{to}^{(s)}, \sigma_{g\alpha\gamma}^2\right)$.

6b. Calculate acceptance probability a.

6c. Set $(\alpha\gamma)_{to}^{(s+1)} \leftarrow y$ with probability a;

otherwise set $(\alpha\gamma)_{to}^{(s+1)} \leftarrow (\alpha\gamma)_{to}^{(s)}$.

Step 7. Repeat steps 7a to 7c for $i = 1$ to n_i.

Repeat steps 7a to 7c for $o = 1$ to n_o.

7a. Draw a sample $y \sim N\left((\beta\gamma)_{io}^{(s)}, \sigma_{g\beta\gamma}^2\right)$.

7b. Calculate acceptance probability a.

7c. Set $(\beta\gamma)_{io}^{(s+1)} \leftarrow y$ with probability a;

otherwise set $(\beta\gamma)_{io}^{(s+1)} \leftarrow (\beta\gamma)_{io}^{(s)}$.

Step 8. Draw a sample $y \sim N\left(\sigma_\alpha^{2(s)}, \sigma_{\sigma_\alpha}^2\right)$

Calculate acceptance probability a.

Set $\sigma_\alpha^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_\alpha^{2(s+1)} \leftarrow \sigma_\alpha^{2(s)}$.

Step 9. Draw a sample $y \sim N\left(\sigma_\beta^{2(s)}, \sigma_{\sigma_\beta}^2\right)$

Calculate acceptance probability a.

Set $\sigma_\beta^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_\beta^{2(s+1)} \leftarrow \sigma_\beta^{2(s)}$.

Step 10. Draw a sample $y \sim N\left(\sigma_\gamma^{2(s)}, \sigma_{\sigma_\gamma}^2\right)$

Calculate acceptance probability a.

Set $\sigma_\gamma^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_\gamma^{2(s+1)} \leftarrow \sigma_\gamma^{2(s)}$.

Step 11. Draw a sample $y \sim N\left(\sigma_{\alpha\beta}^{2(s)}, \sigma_{\sigma_{\alpha\beta}}^2\right)$.

Calculate acceptance probability a.

Set $\sigma_{\alpha\beta}^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_{\alpha\beta}^{2(s+1)} \leftarrow \sigma_{\alpha\beta}^{2(s)}$.

Step 12. Draw a sample $y \sim N\left(\sigma_{\alpha\gamma}^{2(s)}, \sigma_{\sigma_{\alpha\gamma}}^2\right)$.

Calculate acceptance probability a.

Set $\sigma_{\alpha\gamma}^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_{\alpha\gamma}^{2(s+1)} \leftarrow \sigma_{\alpha\gamma}^{2(s)}$.

Step 13. Draw a sample $y \sim N\left(\sigma_{\beta\gamma}^{2(s)}, \sigma_{\sigma_{\beta\gamma}}^2\right)$.

Calculate acceptance probability a.

Set $\sigma_{\beta\gamma}^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_{\beta\gamma}^{2(s+1)} \leftarrow \sigma_{\beta\gamma}^{2(s)}$.

Step 14. Draw a sample $y \sim N\left(\sigma_{e}^{2(s)}, \sigma_{\sigma_{e}}^2\right)$

Calculate acceptance probability a.

Set $\sigma_{e}^{2(s+1)} \leftarrow y$ with probability a; otherwise set $\sigma_{e}^{2(s+1)} \leftarrow \sigma_{e}^{2(s)}$.

Step 15. Set $s \leftarrow s + 1$.

If s does not reach the number set as the total number of iterations in MCMC, return to step 1.

Using samples $\mu_G^{(s)}$, $\alpha_t^{(s)}$, $\beta_i^{(s)}$, $\gamma_o^{(s)}$, $(\alpha\beta)_{ti}^{(s)}$, $(\alpha\gamma)_{to}^{(s)}$, $(\beta\gamma)_{io}^{(s)}$, $\sigma_\alpha^{2(s)}$, $\sigma_\beta^{2(s)}$, $\sigma_\gamma^{2(s)}$, $\sigma_{\alpha\beta}^{2(s)}$, $\sigma_{\alpha\gamma}^{2(s)}$, $\sigma_{\beta\gamma}^{2(s)}$, and $\sigma_e^{2(s)}$, we can estimate various posterior distributions, on which D study will be conducted.

Examples (G study and D study)

With the models and algorithms in the previous section, posterior distributions of variances of random effects can be estimated. Using these estimations, posterior analyses can be conducted. In this section, two hypothetical data sets generated by computer simulations of one-facet $(n_i = 5)$ and two-facet $(n_i = 5, n_o = 3)$ designs were analyzed (G study), and posterior distributions of relative generalizability coefficients were estimated for various designs (D study).

One-facet design

Table 1.Hypothetical data set of one-facet design.

Targtet	Item1	Item2	Item3	Item4	Item5
1	6	5	6	5	3
2	4	4	4	5	6
3	3	1	1	5	3
4	8	5	8	7	5
5	6	2	5	5	3
6	1	3	1	2	2
7	3	2	2	3	4
8	4	3	4	6	3
9	1	3	2	3	2
10	5	3	4	3	3
11	6	4	6	8	6
12	3	3	4	2	3
13	7	3	5	3	4
14	9	6	7	8	8
15	8	3	5	4	4
16	3	4	6	2	3
17	3	2	5	2	4
18	4	1	4	1	4
19	5	6	8	6	5
20	5	4	3	5	7

Table 1 shows hypothetical data of a one-facet design. To estimate posterior distributions of variances σ_α^2, σ_β^2, and σ_e^2 in model (3), the algorithm of the previous section was applied. Initial values of variances were estimated by the classical method of ANOVA (Kirk, 1995; Kutner, Nachtsheim, Neter, and Li, 2005; Myers and Well, 2003). If the classical estimation of σ^2 is not positive (Cardinet, Johnson, and Pini, 2010), σ^2 is set at 0.01. Using these estimates of variances, say σ^2, parameters of $Inv-Gamma(a, b)$ were set as follows:

$$a = 3 + \frac{1}{25}, \quad b = \sigma^2\left(4 + \frac{1}{25}\right).$$

With these values, the mode and standard deviation are given as follows:

$$mode \approx \sigma^2, \qquad standard\ deviation \approx 2\sigma^2.$$

Variances of normal proposal distributions were adjusted in the adaptive MCMC before the main MCMC so that acceptance probabilities are between 0.3 and 0.5 (Rosenthal, 2011). The initial values in the main MCMC were set as corresponding means in the adaptive MCMC, so no burn-in was used. The length of the main MCMC was 10000, so the number of samples from the posterior distribution was 10000. From samples $\sigma_\alpha^{2(s)}$s and $\sigma_e^{2(s)}$; $s = 1, \cdots, 10000$, sample values $\rho_t^{2(s)}$s of the relative generalizability coefficient, which represents reliability with respect to the target's relative position (Cardinet, Johnson, and Pini, 2010; Furr and Bacharach, 2008), were calculated as follows:

$$\rho_t^{2(s)} = \frac{\sigma_\alpha^{2(s)}}{\sigma_\alpha^{2(s)} + \frac{\sigma_e^{2(s)}}{n_i}}. \tag{10}$$

The denominator of (10) contains the components of (3), that have effects on relative positions of targets.

Sample values of generalizability coefficients (10) can be calculated for various values of n_i other than $n_i = 5$, using sample values of $\sigma_\alpha^{2(s)}$s and $\sigma_e^{2(s)}$s from MCMC for $n_i = 5$ (Table 1). These sample values of (10) for various values of n_i comprise sample values of the posterior distribution of generalizability coefficient (10) for respective value of n_i. Using the sample values of the posterior distribution for each value of $n_i = 1$ to 7 , the median and 95% central interval (CI; Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin, 2014) were calculated and shown in Figure 1. Figure 1 shows medians and boundaries of 95% CI for $n_i = 1$ to 7. From medians, we conclude that relative generalizability coefficients are sufficiently large when n_i is at least 3. But when we check 95% CIs, we will decide to choose n_i, which is at least 6.

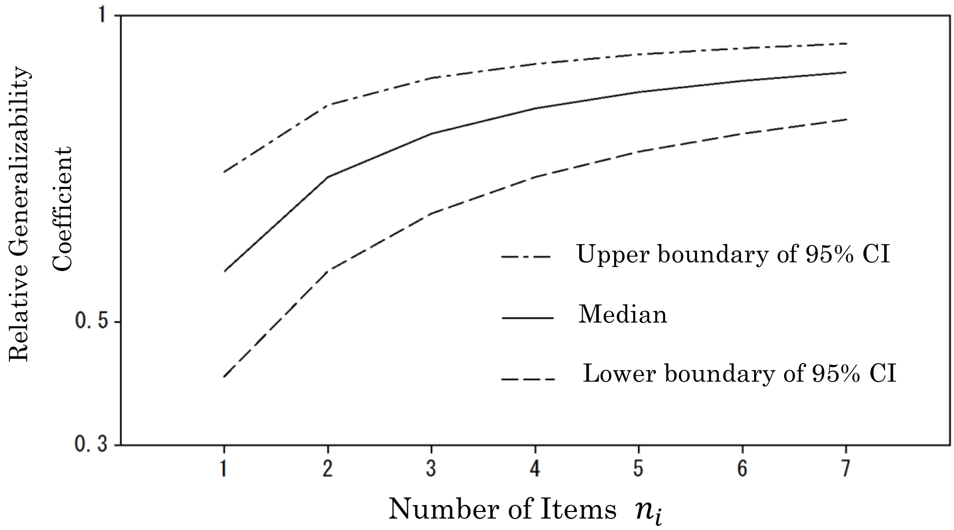

Figure 1. 95% central intervals (CIs) of relative generalizability coefficients for various n_is.

Two-facet design

Table 2 shows hypothetical data of a two-facet design. To estimate posterior distributions of variances in model (8), the algorithm of the previous section was applied. The procedure was basically the same as that in the previous section, except that model (7) was used instead of (2). The length of the main MCMC was 10000, the same length as in the one-facet design. Since initial values of the main MCMC were set as means of the adaptive stage, no burn-in was used, and the number of samples from the posterior distribution was 10000. Sample values $\rho_t^{2(s)}$ s of relative generalizability coefficient were calculated as follows:

$$\rho_t^{2(s)} = \frac{\sigma_\alpha^{2(s)}}{\sigma_\alpha^{2(s)} + \frac{\sigma_{\alpha\beta}^{2(s)}}{n_i} + \frac{\sigma_{\alpha\gamma}^{2(s)}}{n_o} + \frac{\sigma_e^{2(s)}}{n_i n_o}}. \tag{11}$$

The denominator of (11) comprises the components of (8) that have effects on relative positions of targets.

Table 2.　Hypothetical data set of two-facet design.

Target	R1					R2					R3				
	I1	I2	I3	I4	I5	I1	I2	I3	I4	I5	I1	I2	I3	I4	I5
1	5	4	6	6	2	6	5	3	4	5	6	1	6	3	5
2	7	7	5	4	4	6	6	6	6	1	4	3	5	4	4
3	3	2	4	3	3	3	5	5	3	4	3	4	4	2	2
4	10	8	9	7	4	7	6	6	7	4	7	7	10	7	6
5	5	4	5	4	1	7	4	3	4	5	4	3	8	4	4
6	5	2	1	1	1	3	2	1	1	4	1	1	1	2	2
7	3	2	3	2	4	3	4	5	3	3	6	2	1	3	4
8	5	3	5	4	3	7	3	6	6	3	3	3	3	3	4
9	3	2	1	3	1	4	2	1	2	2	4	1	3	3	5
10	5	2	5	5	2	6	1	3	2	2	4	2	4	2	2
11	9	4	4	2	4	9	5	7	7	7	6	3	6	3	5
12	4	2	4	1	1	6	4	2	3	5	4	3	5	1	5
13	6	5	2	5	6	5	1	3	4	3	4	5	6	5	3
14	9	7	6	9	7	9	7	7	10	7	10	8	10	7	8
15	5	4	7	7	8	5	4	4	6	6	5	7	4	7	5
16	7	4	6	5	6	5	4	3	4	6	5	3	3	3	5
16	6	2	4	4	4	3	2	4	5	2	3	2	2	2	3
18	6	1	5	4	1	3	1	1	1	1	6	2	5	1	1
19	8	8	8	6	4	8	6	7	7	7	6	8	7	7	8
20	8	5	8	5	8	5	5	7	4	4	4	7	6	1	5

Sample values of generalizability coefficients (11) can be calculated for various values of n_i and n_o other than $n_i = 5$ and $n_o = 3$, using sample values of $\sigma_\alpha^{2(s)}$s, $\sigma_{\alpha\beta}^{2(s)}$s, $\sigma_{\alpha\gamma}^{2(s)}$s, and $\sigma_e^{2(s)}$s from MCMC for $n_i = 5$ and $n_o = 3$ (Table 2). These sample values of (11) for various values of n_i and n_o comprise sample values of the posterior distribution of generalizability coefficient (11) for respective values of n_i and n_o . Using the sample values of the posterior distribution for each value of $n_i = 1$ to 7 with fixed value $n_o = 5$, the median and 95% central intervals (CIs) were calculated and shown in Figure 2. When the number of observers is five $(n_o = 5)$, considering medians, we conclude that any number of items is sufficient. However, lower boundaries of 95% CIs indicate that at least two items are required for a test to be statistically reliable.

Figure 3 shows 95% CIs of relative generalizability coefficients for various n_os with $n_i = 5$.When the number of items is five $(n_i = 5)$, any number of observers is sufficient considering medians. However, lower boundaries of 95% CIs indicate that at least two observers are required for a test to be statistically reliable.

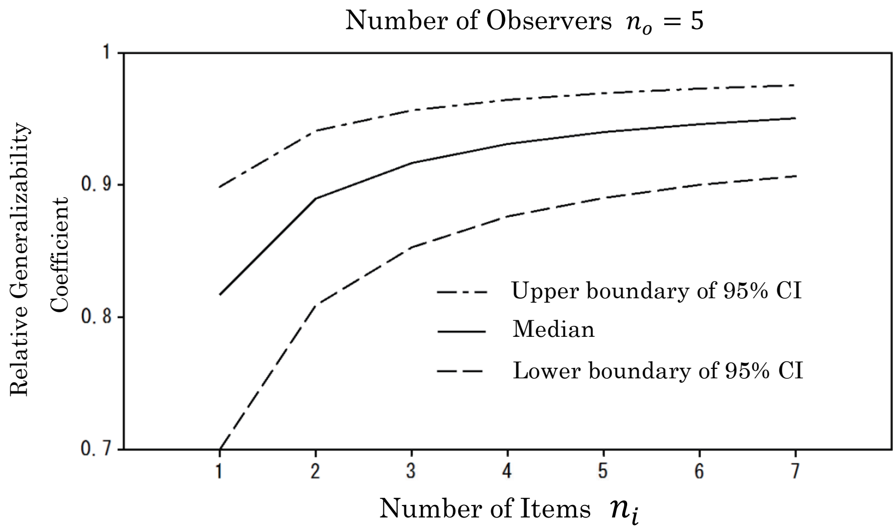

Figure 2. 95% central intervals (CIs) of relative generalizability coefficients for various n_is with $n_0 = 5$.

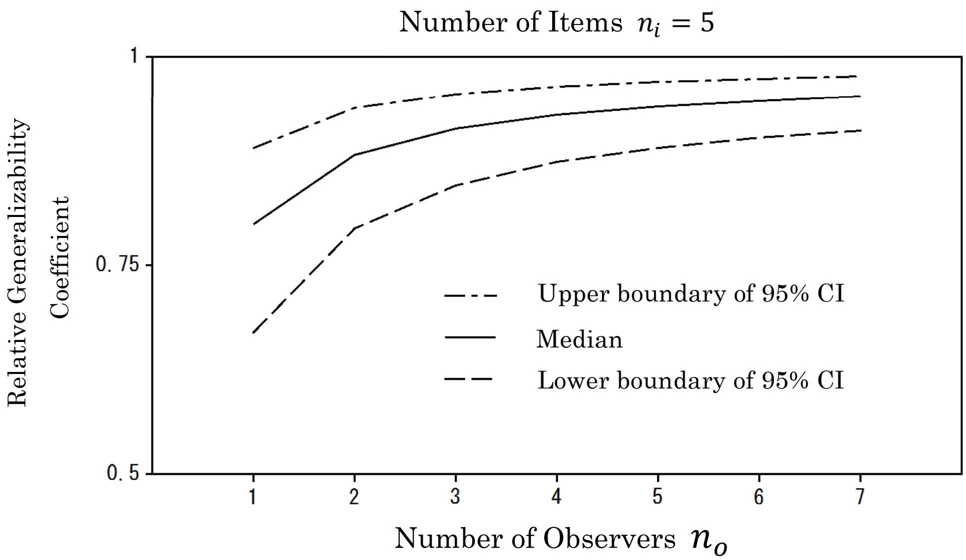

Figure 3. 95% central intervals (CIs) of relative generalizability coefficients for various n_os with $n_i = 5$.

Discussion

The framework of generalizability theory, G study and D study, corresponds well to that of Bayesian analysis, setting up a model, conditioning on a data, and evaluation. To apply the Bayesian approach, this study employed stochastic models, in that random effects are represented as hierarchical prior distributions. Decompositions of variances are derived using well-known rules of probability theory. With these decompositions, generalizability coefficients can be calculated.

The hypothetical data sets of one-facet and two-facet designs were analyzed according to the proposed algorithms. To conduct Bayesian analysis by the proposed models, Metropolis-within-Gibbs algorithms were employed. Posterior distributions of variances were successfully estimated, using inverse gamma distributions as prior distributions for variances. Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin (2014) state that in most problems, using a weakly informative prior distribution that includes a small amount of real-world information is preferable. In this study, parameters of the inverse-gamma distributions were set using estimates by ANOVA. With sample values from posterior distributions of variances, posterior distributions of relative generalizability coefficients were estimated, and 95% CIs were derived. 95% CIs provided more information than point estimations, i.e., medians. It should be emphasized that although estimates of variances by ANOVA can be negative (Cardinet, Johnson, and Pini, 2010), the proposed methods always provide nonnegative estimates by virtue of prior distributions.

In this study, only relative generalizability coefficients were treated. Obviously, an absolute generalizability coefficient can be estimated in the same way as a relative generalizability coefficient

References

Alcalá-Quintana, R., and García-Pérez, M. A. (2004). The role of parametric assumptions in adaptive Bayesian estimation. *Psychological Methods*, 9, 250-271.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Brennan, R. L. (2011). Generalizability theory and classical test theory. Applied *Measurement in Education*, 2011, 24, 1-21.

Cardinet, J., Johnson, S., and Pini, G. (2010). *Applying generalizability theory using EduG*. New York: Routledge.

Carlin, B. P., and Louis, T. A. (2009). *Bayesian methods for data analysis, 3rd ed*. Boca Raton: Chapman & Hall/CRC.

Furr, R. M., and Bacharach, V. R. (2008). *Psychometrics: An introduction*. Los Angeles: SAGE Publications.

Gamerman, D., and Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference, 2nd ed*. New York: Chapman & Hall/CRC.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis*, 1, 515-534.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*, ed. Boca Raton: CRC Press.

Glas, C. A. W. (2012). Generalizability theory and item response theory. In *Psychometrics in practice at RCEC*, 2012, 1-13. Retrieved from http://dx.doi.org/10.3990/3.9789036533744.ch1

Hogg, R. V., McKean, J. W., and Craig, A. T. (2005). *Introduction to mathematical statistics, 6th ed*. Upper Saddle River: Pearson Prentice Hall.

Kingdom, F. A. A., and Prins, N. (2010). Psychophysics: *A practical introduction*. San Diego: Academic Press.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences, 3rd ed*. Pacific Grove: Brooks/Cole

Publishing Company.

Kreiter, C. D. (2010). Using generalizability theory to evaluate the applicability of a serial Bayes model in estimating the positive predictive value of multiple psychological or medical tests. *Psychology*, 2010, 1, 194-198.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Amsterdam: Elsevier.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied linear statistical models, 5th ed.* Boston: McGraw Hill.

LoPilato, A. C., Carter, N. T., and Wang, M. (2014). Updating generalizability theory in management research: Bayesian estimation of variance components. *Journal of Management*, 2014, October, 1-26. Retrieved from http://jom.sagepub.com/content/early/2014/10/08/0149206314554215

Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2013). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton: CRC Press.

Myers, J. L. and Well, A. D. (2003). *Research design and statistical analysis, 2nd ed*. Mahwah: Lawrence Erlbaum Associates, Publishers.

Okamoto, Y. (2013). A direct Bayesian estimation of reliability. *Behaviormetrika*, 40, 149-168.

Robert, C. P., and Casella, G. (2010a). *Introducing Monte Carlo methods with R*. New York: Springer.

Robert, C. P., and Casella, G. (2010b). *Monte Carlo statistical methods*. New York: Springer.

Rosenthal, J. S. (2011). Optimal proposal distributions and adaptive MCMC. In S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng (Eds.), *Handbook of Markov chain Monte Carlo* (pp. 93-111). Boca Raton: CRC Press.

Shavelson, R. J. and Webb, N. M. (1991). *Generalizability theory*. Newbury Park: SAGE Publications.

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