Paper SD-130

Impact of Q-matrix Misspecification on Cognitive Attribute Estimation in the Crossed Random Effects LLTM Model with the SAS® GLIMMIX Procedure

Yi-Hsin Chen, University of South Florida; Isaac Y. Li, University of South Florida; Chunhua Cao, University of South Florida; Yan Wang, University of South Florida

ABSTRACT

A simulation study is designed to explore the impact of Q-matrix misspecification on parameter estimation in the crossed random effects linear logistic test model using the SAS® GLIMMIX procedure. In addition, the impact of the interactions of Q-matrix misspecification with other manipulated factors such as population distribution, sample size, and Q-matrix density on parameter estimation are investigated as well. Datasets are simulated based on the model of interest using the SAS/IML package. The population distributions include normal, negatively-skewed, and positively-skewed distributions. The sample sizes are 50, 250, and 500. The percentages of misspecification in the Q-matrix are 2.4%, 4.8%, and 9.6%. Three types of misspecification are over-, under-, and balanced-misspecifications. For each condition, 1000 replications are generated. Cognitive attribute parameters are estimated by applying the SAS GLIMMIX procedure. The parameters of interest in this study are the estimates for cognitive attributes. The true cognitive attribute parameters for the sparse Q-matrix are η_1 = 2.152, η_2 = 1.229, η_3 = -.468, η_4 = 1.907, η_5 = 1.051, η_6 = .086, η_7 = .141, and η_8 = -.474. The sparse Q-matrix has only 48 out of 168 entries that contain 1s (approximately 30% 1s). In contrast, the dense Q-matrix has 96 out of 168 entries that contain 1s (approximately 60% 1s). Misspecified entries in the Q-matrix are randomly assigned using SAS based on design factors of misspecification percent and type as mentioned earlier. The results indicate that misspecification type and percent have a considerable impact on the bias and root mean squared error of attribute estimates, especially under the conditions of the high percent misspecification and of the over-misspecification. However, attribute correlation between the estimated and true parameters is not affected by misspecification type and percent. Since the Q-matrix is an indispensable element in applying the crossed random effects linear logistic test model, specifying an appropriate Q-matrix, therefore, is a crucial task and must be completed with generous assistance from content and subject experts.

Keywords: CROSS RANDOM EFFECTS; LLTM; SIMULATION; Q-MATRIX MISSPECIFICATION; SAS GLIMMIX

INTRODUCTION

The linear logistic test model (LLTM) proposed by Fisch (1973) has been applied widely for various purposes (e.g., item selection, test development, construct validation) in educational contexts because this model allows one to decompose item difficulty into cognitive components in terms of test tasks. One of the important features of applying the LLTM is to identify a list of cognitive components related to test tasks and construct the relations between test tasks and cognitive components before estimating the parameters. The final product is called a design matrix or so-called a Q-matrix that is presented as an *I* (number of items) x *K* (number of cognitive components) matrix. Examples of the Q-matrices can be found in Tables 1 and 2 in the method section. The Q-matrix can be a binary matrix where 1s indicate the presence of cognitive components on particular items, otherwise 0.

The LLTM is often considered an extension of the Rasch model because the mathematical equation of the LLTM can be expanded directly from the Rasch model (e.g., Chen, MacDonald, & Leu, 2011; Hartig, Frey, Nold, Klieme, 2012) and all the features of the Rasch model (e.g., one-parameter model, specific objectivity, sufficient statistic, parameter separability) can be applied to the LLTM as well. The mathematical equation of the Rasch model is presented as follows:

$$P(X_{ip}=1 \mid \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1+\exp(\theta_p - \beta_i)},$$

where $P(X_{ip}=1)$ represents the probability of the correct response to item *i* for person *p*, θ_p is the ability level of person *p*, and β_i is the item difficulty of item *i*. In the LLTM, a linear combination of cognitive components replaces item difficulty, β_i, of the Rasch model. In other words, the item parameters of the Rasch model are replaced by a product of cognitive components and their weights (i.e., entries in a Q-matrix), which can be conceptualized as

$$\hat{\beta}_i = \sum_{k=1}^{K} q_{ik}\eta_k + c,$$

where q_{ik}, an entry in a Q-matrix, is the fixed and predetermined weight assgined to cognitive component *k* that is involved in item *I,* and η_k is the estimated parameter for cognitive component *k*, and *c* is the normalizing constant and is simply defined as the mean of the β_i estimates under the Rasch model. Thus, the mathematical equation of the

LLTM is presented as follows:

$$P(X_{ikp}=1\,|\,\theta_j,q_{ik},\eta_k)=\frac{\exp(\theta_p-\sum_{k=1}^{K}q_{ik}\eta_k+c)}{1+\exp(\theta_p-\sum_{k=1}^{K}q_{ik}\eta_k+c)}\,.$$

A well-known assumption behind the LLTM is that the variance of item difficulties is explained completely by cognitive components (i.e., $\beta_i=\hat{\beta}_i$). This assumption does not take into account the fact of sampling items from an item population that results in item sampling variance (De Boeck, 2008). In applications, the LLTM assumption leads to the same item difficulty estimates for the items requiring the same cognitive components and the same weights. This may not always be the case. Thus, this can be considered a disadvantage of applying the LLTM for the purpose of decomposing item difficulty into cognitive components (e.g., Hartig, Frey, Nold, & Klieme, 2012; Janssen, Schepers, & Peres, 2004).) To overcome the disadvantage of the LLTM, Janssen and colleagues (2003, 2004) proposed the crossed random-effects logistic linear test model (CRELLTM), just like in a regular regression model by adding the error term on item difficulty to relax the assumption of the LLTM as the equation below where ε_i is an error term with a normal distribution, $\varepsilon_i \approx N(0,\sigma_\varepsilon^2)$.

$$\beta_i=\hat{\beta}_i+\varepsilon_i=(\sum_{k=1}^{K}q_{ik}\eta_k+c)+\varepsilon_i$$

FORMULATION AND ESTIMATION OF THE CRELLTM USING THE GLIMMIX PROCEDURE

In SAS, the NLMIXED procedure has been used widely to formulate diverse item response theory (IRT) models because of its availability of software and flexibility of modeling (De Boeck & Wilson, 2004; Wang & Jin, 2010). However, the SAS NLMIXED procedure cannot be used in this study because random effects are required for person ability and item difficulty simultaneously, which is called crossed random effects (De Boeck & Wilson, 2004). The newly-developed GLIMMIX procedure in SAS is applicable for the models with the cross random effects (Wang & Jin, 2010). Historically, the GLIMMIX procedure could be applied from a SAS macro as an add-on product in SAS 9.1. Now as an individual package, a lot of improvements have been made in SAS 9.3 (see Li, Chen, & Kromrey, 2013). Like the NLMIXED, the GLIMMIX performs estimation and statistical inference for generalized linear mixed models (GLMMs) that extends the class of generalized linear models (GLMs) by incorporating normally distributed random effects. The GLIMMIX procedure has been applied to formulate few IRT-related models compared to the NLMIXED and there has been only one study that examined its efficiency in parameter recovery (Cao, Wang, Chen, & Li, 2014).

Using generalized linear or nonlinear models for formulation of IRT models, such as the CRELLTM studied in this paper, there are the three required specifications, including (1) the random or distribution component, (2) the systematic component, and (3) the link component (De Boeck & Wilson, 2004; Wang & Jin, 2010). The random or distribution component specifies the distribution of the data. For the dichotomous or binary data, it is appropriate to assign the Bernoulli or binary distribution. The Bernoulli distribution is one of the exponential distributions. The distribution component describes the relation between the distribution of the data (Y_{ip}) and the expected value of the distribution (μ_{ip}), which is the probability of correct response (P_{ip}) for the dichotomous data. The systematic component defines a linear or nonlinear function of the predictors for persons and items, denoted η_{ip}. The equation of the systematic component for the CRELLTM is shown below.

$$\eta_{ikp}=\theta_p-\beta_i=\theta_p-[(\sum_{k=1}^{K}q_{ik}\eta_k+c)+\varepsilon_i]$$

$$\theta_p \approx N(\mu_\theta,\sigma_\theta^2)\text{ and }\varepsilon_i \approx N(0,\sigma_\varepsilon^2)$$

Like in the Rasch model, the mean of item difficulty parameters is set to be 0 and the mean of person ability parameters is freely estimated in the CRELLTM. The link component connects the expected value of the data to η_{ip}, which is the systematic, nonlinear function for the CRELLTM. The logit link function is a more commonly used link function than the probit link function for IRT models. The link function of the CRELLTM can be presented as follows:

$$\log(\frac{p_{ikp}}{1-P_{ikp}})=\eta_{ikp}=\theta_p-[(\sum_{k=1}^{K}q_{ik}\eta_k+c)+\varepsilon_i]$$

By conducting mathematical transformation, the probability of correct response based on ability level, cognitive component, and the Q-matrix for the CRELLTM can be expressed below and all the notations can be found in the previous paragraphs.

$$P(X_{ikp}=1\,|\,\theta_p,q_{ik},\eta_k)=\frac{\exp(\theta_p-[(\sum_{k=1}^{K}q_{ik}\eta_k+c)+\varepsilon_i]}{1+\exp(\theta_p-[(\sum_{k=1}^{K}q_{ik}\eta_k+c)+\varepsilon_i]}$$

There are two categories of estimation methods under the GLIMMIX procedure: (1) pseudo-likelihood under linearization and (2) maximum likelihood with Laplace approximation (METHOD=LAPLACE) or adaptive quadrature (METHOD=QUAD). Pseudo-likelihood estimation methods tend to yield biased estimates for non-normal data with small sample sizes. Maximum likelihood estimation with Laplace integral approximation seems to an appropriate estimation method for the CRELLTM because the CRELLTM is a random-person random-item model that does not require the fixed effect (R-side effect in SAS) in parameter estimation processes (SAS, 2011). Please refer to *SAS User's Guide* (2011) for the detailed information of these estimation methods.

Below is the SAS code of the CRELLTM to obtain cognitive attribute estimates using the SAS GLIMMIX procedure. As for the data import and the Q-matrix specification for the CRELLTM, please refer to Cao, Wang, Chen, and Li (2014). In the code, resp1 represents the dependent variable and there are four cognitive components, a1 to a4.

```
PROC GLIMMIX data = combine method = laplace;
CLASS item person;
MODEL resp1 (descending)=a1 a2 a3 a4/s noint link=logit dist=binary error=binomial;
RANDOM int / subject=person s;
RANDOM int / subject=item s;
run;
```

RESEARCH PURPOSE AND SPECIFIC QUESTIONS

For the purpose of using the LLTM and LLTM with random item effects models, it is crucial to identify well-defined cognitive attributes (or task characteristics) and establish the appropriate relationships between test tasks and cognitive components or attributes (i.e., the Q-matrix) in addition to determining the psychometric model for the probability of the correct responses. For the LLTM without random effects, studies have examined if the model provides accurate parameter estimates for cognitive attributes (e.g., Cassuto, 1996; Green & Smith, 1987; MacDonald, 2014) and if the model is sensitive to misspecification of the Q-matrix (e.g., Baker, 1993). For the LLTM with random item effects, its sensitivity to misspecification of the Q-matrix has not been examined yet, to our knowledge, at least using the SAS GLIMMIX procedure as an estimation tool.

Based on the aforementioned rationale, a series of simulations were conducted in this study to explore the effects of Q-matrix misspecification on parameter estimation in the LLTM model with random item effects. The SAS GLIMMIX procedure was utilized to estimate parameters of cognitive attributes. Q-matrix misspecification involves misspecification percent (i.e., 2.4%, 4.8%, and 9.6%) and misspecification type (i.e., over-, under-, and balanced-misspecifications). In addition, the impact of population distribution, sample size, and Q-matrix density with Q-matrix misspecification on parameter estimation was also explored. There were five specific research questions: (a) How does Q-matrix misspecification percent affect parameter estimation? (b) How does Q-matrix misspecification type affect parameter estimation? (c) Are the effects of Q-matrix misspecification on parameter estimation different across various populations? (d) Are the effects of Q-matrix misspecification on parameter estimation different across sample sizes? (e) Do the effects of misspecification on parameter estimation vary between dense and sparse Q-matrices?

METHOD

DATA GENERATION

Datasets were simulated based on the LLTM model with random item effects using the SAS/IML package. The population distributions included normal (sk=0, kur=0), negatively-skewed (sk=-0.5, kur=3), and positively-skewed (sk=0.5, kur=3) distributions. The sample sizes were 50, 250, and 1000. The percentages of misspecification in the Q-matrix were 2.4% (4 entries), 4.8% (8 entries), and 9.6% (16 entries). Three types of misspecification were over-misspecification (0s→1s), under-misspecification (1s→0s), and balanced-misspecification (0s→1s and 1s→0s). For each condition, 1000 replications were generated. Parameters in the LLTM with random item effects were estimated by applying the SAS GLIMMIX procedure. The parameters of interest in this study were the estimates for cognitive attributes.

Q-MATRICES

Like Baker's (1993) study, the sparse and dense Q-matrices that were extracted from Fischer and Formman (1972) and Medina-Diaz (1993), respectively, were used in this study. The true cognitive attribute parameters for the sparse Q-matrix were η_1 = 2.152, η_2 = 1.229, η_3 = -.468, η_4 = 1.907, η_5 = 1.051, η_6 = .086, η_7 = .141, and η_8 = -.474. In the sparse Q-matrix, a total of 21 items with 8 cognitive attributes was involved in the original sparse Q-matrix in Baker's (1992) study. As seen in Table 1, the sparse Q-matrix had only 48 out of 168 entries that contained 1s (approximately

30% 1s). In contrast, the dense Q-matrix had 96 out of 168 entries that contained 1s (approximately 60% 1s) as shown in Table 2. The Q-matrices shown in Tables 1 and 2 represent the true Q-matrices. Misspecified entries in the Q-matrix were randomly assigned using SAS based on design factors of misspecification percent and type mentioned in the previous paragraph. For instance, under the condition of 2.4% under-misspecified Q-matrix, the SAS program was designed to randomly change 4 entries in the true Q-matrix from 1s to 0s.

				Cognitive Attribute					
Item	1	2	3	4	5	6	7	8	Total
1	0	1	0	0	0	0	1	0	2
2	1	0	0	1	0	0	1	0	3
3	0	0	0	0	0	1	1	0	2
4	1	1	0	0	0	0	0	1	3
5	0	1	0	0	0	0	0	1	2
6	0	0	0	0	1	0	0	1	2
7	0	0	1	1	0	0	0	1	3
8	0	1	0	0	0	0	1	0	2
9	1	0	0	0	0	0	1	0	2
10	1	0	1	1	0	0	0	0	3
11	0	0	1	1	0	0	0	0	2
12	0	0	0	0	1	1	0	0	1
13	0	0	1	0	0	0	1	0	2
14	0	0	0	0	1	1	0	1	3
15	0	0	0	1	0	0	0	1	2
16	0	0	0	0	1	0	1	0	2
17	0	1	0	0	0	0	1	0	1
18	0	0	1	0	0	1	1	0	3
19	0	0	0	0	0	1	0	1	2
20	1	0	0	0	0	0	0	1	2
21	0	0	0	0	1	0	0	1	2
Total	5	5	5	5	5	5	9	9	48

Table 1. The Sparse Q-matrix

				Cognitive Attribute					
Item	1	2	3	4	5	6	7	8	Total
1	1	0	1	0	1	1	0	0	4
2	0	0	0	1	0	1	0	1	3
3	0	1	1	1	0	0	1	0	4
4	1	1	0	0	1	0	0	0	3
5	0	0	1	0	0	1	1	1	4
6	0	1	0	1	0	1	0	1	4
7	1	0	1	0	1	0	1	0	4
8	1	0	0	1	0	1	1	0	4
9	0	1	1	0	1	1	0	1	5
10	1	1	0	1	0	1	0	1	5
11	1	1	0	1	0	1	0	0	4
12	0	0	1	1	1	0	1	0	4
13	0	1	1	0	1	0	1	1	5
14	1	0	1	1	0	1	0	1	5
15	1	1	0	1	1	0	1	0	5
16	0	1	1	1	0	1	1	1	6
17	1	1	0	1	1	0	1	1	6
18	1	1	0	1	1	1	0	1	6
19	1	0	1	1	1	1	1	0	6
20	1	1	1	0	1	0	1	1	6
21	1	1	1	0	1	1	1	1	7
Total	12	12	12	12	12	12	12	12	96

Table 2. The Dense Q-matrix

EVALUATION CRITERIA

Three decision criteria were used to assess sensitivity of the LLTM with random item effect to Q-matrix misspecification, including bias, root mean square error (RMSE), and correlation. The estimation bias was computed as the average difference between the estimated and true parameters. The formula for estimation bias for cognitive attributes is as follows:

$$Bias = \frac{\sum \hat{\eta} - \eta}{n_{replication}}$$

The RMSE is the square root of the average squared difference between the estimated and true parameters and was used to detect the magnitude of estimation error. The RMSE formula is as follows:

$$RMSE = \sqrt{\frac{\sum (\hat{\eta} - \eta)^2}{n_{replication}}}$$

Person product-moment correlation was used to detect the consistency between the estimated and true sets of parameters. High correlation coefficients indicate that the set of estimated parameters is consistent with the true parameters. Finally, factorial ANOVA analyses with the generalized eta-squared effect size were used to examine what manipulated factors affect bias, RMSE, and correlation. The Cohen's moderate effect size of .0588 was applied as the practical significant level.

RESULTS

To explore the effects of Q-matrix misspecification including type and percent on parameter estimation of cognitive attributes, the boxplots that describe the distribution of bias, RMSE, and correlation for Q-matrix misspecification type and percent were examined. In addition, the eta-squared effect sizes (η^2) of the main effects and the first level interactions with Q-matrix misspecification that were associated with manipulated factors in this study (i.e., sample size, population shape, and Q-matrix density) were computed. The graph for the significant interaction effect was shown as well.

ESTIMATED BIAS

The results of factorial ANOVA analyses with generalized eta-squared effect sizes for estimated bias as shown in Table 3 indicated that the main effect of misspecification types ($\eta2$ = .3064) and the interaction between type and percent of misspecification ($\eta2$ = .0743) were significantly associated with the bias of cognitive attribute estimation using the .0588 as the practical significant level. The main effect of misspecification percent ($\eta2$ = .0014) did not emerge a significant impact on estimated bias. Population shape, Q-matrix density, and sample size did not show interactive impact with Q-matrix misspecification (i.e., type and percent) on estimated bias either.

	Cognitive Attribute
Effect	η^2
Misspecification Type	**0.3064**
Misspecification Type * Misspecification Percent	**0.0743**
Misspecification Type * Shape	0.0430
Misspecification Type * Q-Matrix Density	0.0302
Misspecification Percent * Sample Size	0.0097
Misspecification Percent * Shape	0.0201
Misspecification Type * Sample Size	0.0030
Misspecification Percent	0.0014
Misspecification Percent * Q-Matrix Density	0.0007

Table 3. Effect of misspecification factors on estimated bias of cognitive attribute
Note. Effect sizes are presented in descending order and for significant effects appear in bold.

The most significant impact of manipulated factors on estimated bias of cognitive attributes was misspecification type involving under misspecification, over misspecification, and balanced misspecification. The distributions of the estimated bias are shown in Figure 1. As shown Figure 1, the average estimated bias for the true Q-matrix (M = -.0131) was negligible and close to 0. When the Q-matrix was under-misspecified, parameter estimates seemed to yield positive bias (M = .0607); that is, cognitive attribute parameters were over-estimated when the Q-matrix was misspecified from 1s to 0s. In contrast, there were larger negative bias (M = -.1180) when the Q-matrix was over-misspecified (i.e., from 0s to 1s), compared to under-misspecified Q-matrix. In other words, cognitive attribute parameters tended to be under-estimated when there were many the Q-matrix entries misspecified from 0s to 1s. As for the balanced-misspecified Q-matrix, there was a very small negative bias (M = -.0348). The standard deviations of estimated bias for three types of misspecification and no misspecification were similar, approximately equal to .10.

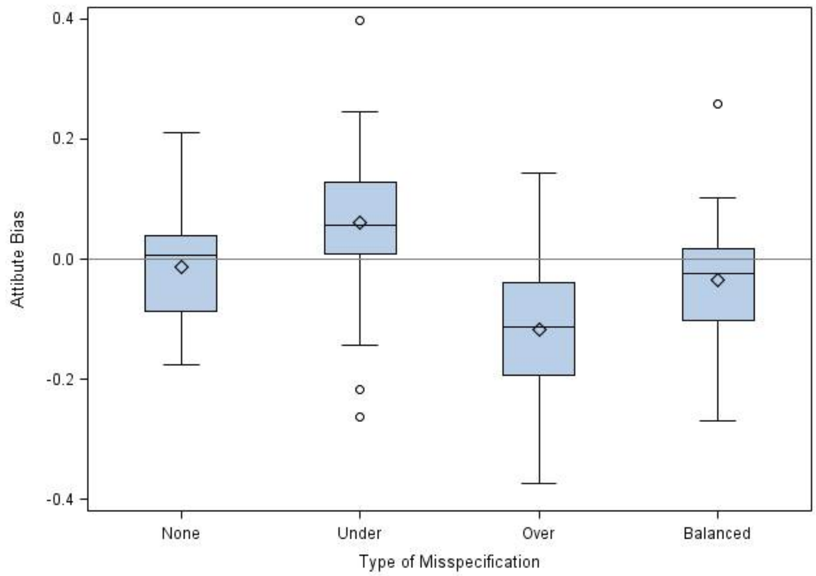

Figure 1. Distributions of estimated bias for cognitive attribute by misspecification type

Figure 2 shows the distribution of estimated bias for cognitive attribute under the no misspecification and three different misspecification rates. As seen in Figure 2, estimated bias slightly increased as percent of misspecification increased, but the mean bias differences among different percentages of misspecification were negligible as evident in the reported eta-squared effect size see Table 3). The means and standard deviations for 0%, 2.4%, 4.8%, and 9.6% of the Q-matrix entry misspecification are -.013 and .103, -.026 and 0.108, -.028 and .119, as well as -.038 and .155, respectively.

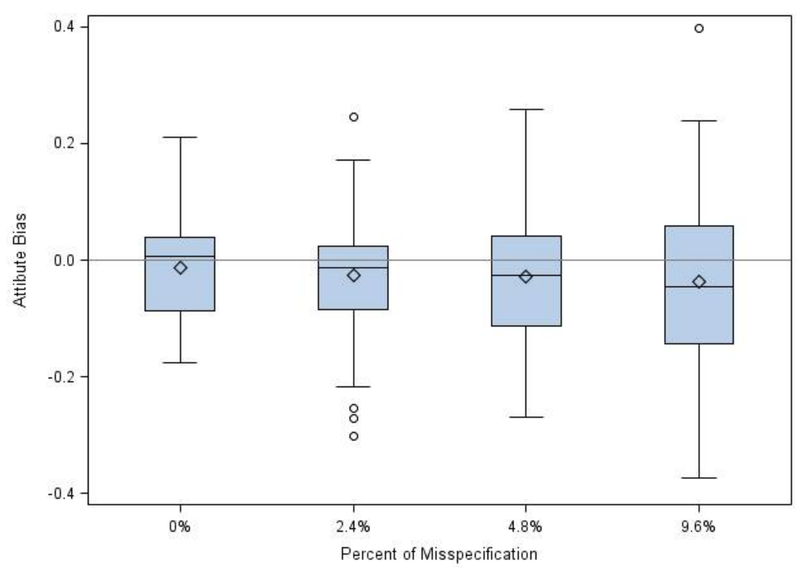

Figure 2. Distributions of estimated bias for cognitive attribute by misspecification percent

Figure 3 shows the significant interaction between misspecification type and percent on estimated bias (η2 = .0743). As shown in the graph, the over-misspecification increased more bias in the negative direction as the misspecification percent increased. The under-misspecification also increased more bias but in the positive direction as the misspecification percent increased. Compared to the over misspecification, the increases in bias for the under misspecification were less. Interestingly, the balanced-misspecification yielded slight increase in estimated bias as the misspecification percent increased.

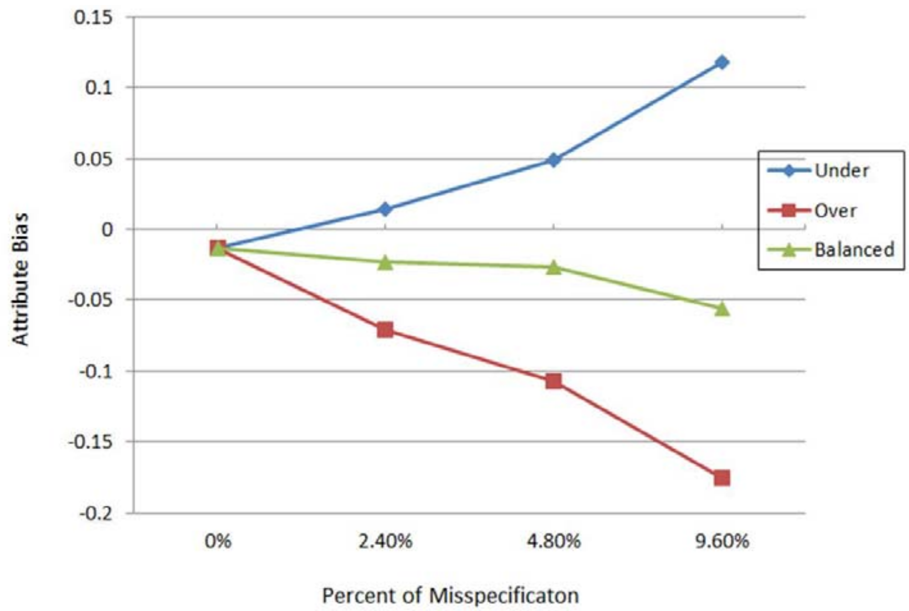

Figure 3. Mean bias by misspecification type and percent

ROOT MEAN SQUARE ERROR

Table 4 shows eta-squared effect sizes of misspecification factors and interactions with other manipulated factors (i.e., sample size, population shape, and Q-matrix density) for RMSE. The results indicated that the main effects of misspecification percent and type were highly associated with RMSE of attribute estimates (both η^2 > .0588). Misspecification factors did not show interactions with other manipulated factors in terms of estimated RMSE. That is, these other factors such as sample size did not affect the impact of Q-matrix misspecification on RMSE of attribute estimates.

Effect	η^2
Misspecification Percent	**0.19280**
Misspecification Type	**0.15755**
Misspecification Type * Sample Size	0.02140
Misspecification Percent * Shape	0.01511
Misspecification Percent * Q-matrix Density	0.01478
Misspecification Type * Misspecification Percent	0.01229
Misspecification Percent * Sample Size	0.00786
Misspecification Type * Q-Matrix Density	0.00768
Misspecification Type * Shape	0.00474

Table 4. Effect of Misspecification Factors on Estimated RMSE of Cognitive Attribute
Note. Effect sizes are presented in descending order and for significant effects appear in bold.

Figure 4 displays the distributions of RMSE for cognitive attribute by misspecification percent. The graph showed that the RMSE significantly increased when the percent of misspecification increased. The RMSE means for 0%, 2.4%, 4.8%, and 9.6% of misspecification in the Q-matrix were .54, .66, .75, and .85, respectively. Figure 5 presents the distributions of RMSE for cognitive attribute by misspecification type. The RMSE means for none, under, over, and balanced misspecifications of the Q-matrix were, .54, .77, .73, and .76, respectively. Larger RMSEs were yielded by three types of misspecification, compared to no misspecification. However, there seemed to be no differences between the means of RMSE of the three types of misspecification.

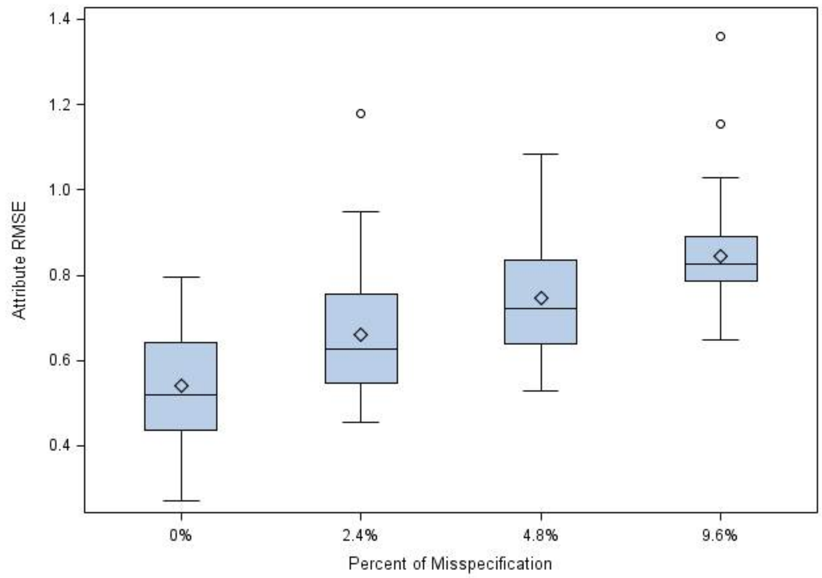

Figure 4. Distributions of RMSE for cognitive attribute by misspecification percent

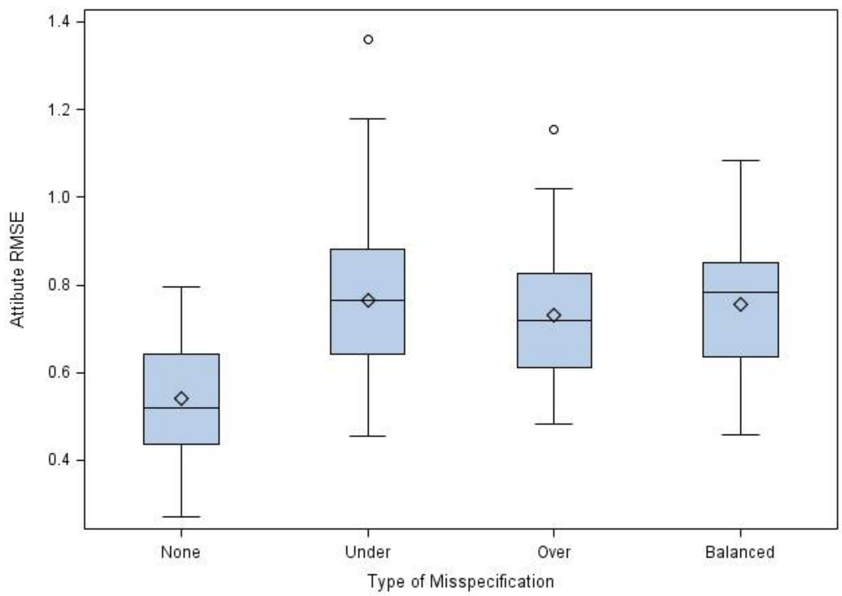

Figure 5. Distributions of RMSE for cognitive attribute by misspecification type

CORRELATION BETWEEN ESTIMATED AND TRUE ATTRIBUTE PARAMETERS

The eta-squared effect sizes of misspecification factors and their interactions with other manipulated factors are reported in Table 5. Unlike attribute bias and RMSE, misspecification factors (i.e., percent and type) showed small impact on attribute correlations between estimated and true attribute parameters (η^2 = .0131 and .0066, respectively). The effects of interactions between misspecification factors and manipulated factors on attribute correlation did not reach the cutoff value of eta-squared effect size (η^2 = .0588).

Cognitive Attribute Effect	η^2
Misspecification Percent * Sample Size	0.0387
Misspecification Type * Sample Size	0.0355
Misspecification Type * Misspecification Percent	0.0284
Misspecification Type * Shape	0.0262
Misspecification Percent	0.0131
Misspecification Type * Q-Matrix Density	0.0126
Misspecification Percent * Q-Matrix Density	0.0113
Misspecification Percent * Shape	0.0107
Misspecification Type	0.0066

Table 5. Effect of misspecification factors on correlation between estimated and true attribute parameters.
Note. Effect sizes are presented in descending order and for significant effects appear in bold.

The mean correlations for 0%, 2.4%, 4.8%, and 9.6% of the Q-matrix misspecification were .86, .87, .88, and .89, respectively. Figure 6 displays the distributions of correlations for no misspecification and three levels of misspecification, indicating no effects of misspecification percent on attribute correlations. Similarly, the mean correlations for zero, under, over, and balanced misspecifications were .86, .88, .89, and .88, respectively. As seen in Figure 7, the correlation distributions for the four misspecification types were similar.

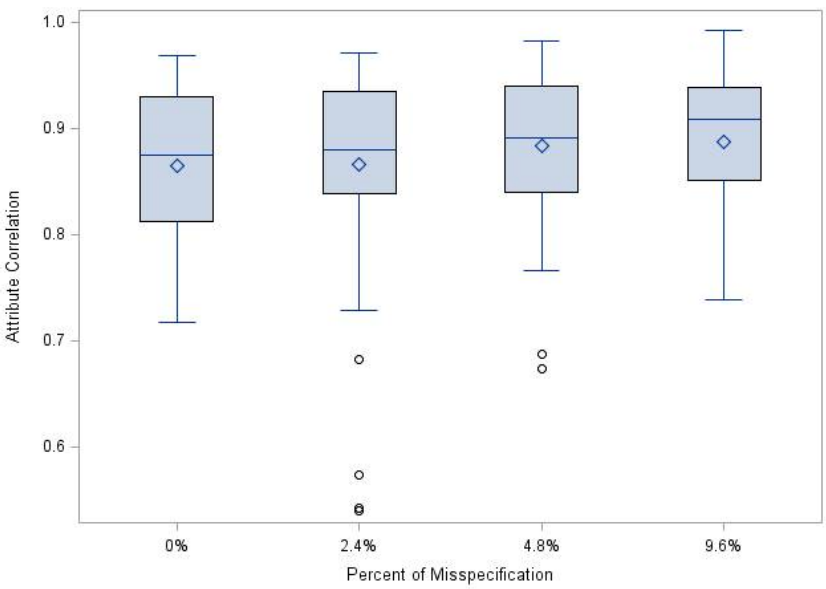

Figure 6. Distributions of correlations for cognitive attribute by misspecification percent

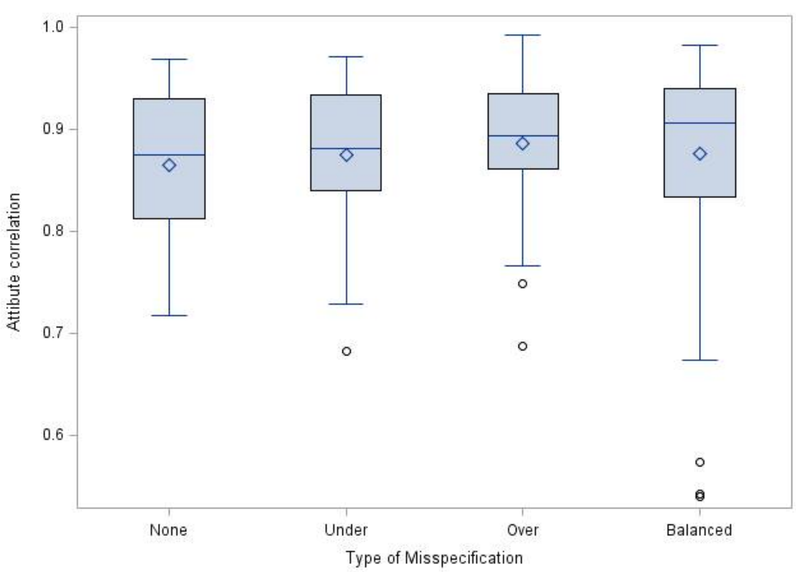

Figure 7. Distributions of correlations for cognitive attribute by misspecification type

CONCLUSION

The LLTM has been a well-known method of estimating cognitive attribute parameters for the purpose of decomposing item difficulty. As mentioned earlier, the LLTM with random item effects, also called the crossed random effects LLTM (CRELLTM; Cao, Wang, Chen, & Li, 2014) was developed to overcome one of the disadvantages of the LLTM, the assumption of the variances of item difficulty being completely accounted for by cognitive attributes. The accuracy and precision of the CRELLTM deployed with the SAS GLIMMIX procedure have been explored by Cao and colleagues (Cao, et al., 2014). However, the sensitivity of this model to misspecification of the Q-matrix in terms of cognitive attribute estimates using the SAS GLIMMIX procedure has not been examined yet. The purpose of this study was intended to provide practitioners and researchers alike with insight into this model and utility of the SAS GLIMMIX procedure.

The results indicated that as the misspecification percent in the Q-matrix increased, the impacts of both over misspecification and under misspecification on attribute estimate bias increased substantially. Interestingly, larger bias was yielded by the over-misspecification in a negative way whereas smaller bias was given by the under-misspecification in a positive way. In other words, with the over-misspecified Q-matrix (i.e., 0s were misspecified by 1s), attributes tended to be under-estimated. In contrast, with the under misspecified Q-matrix (i.e., 1s were misspecified by 0s), attributes were over-estimated. The Q-matrix with the over-misspecification and with higher misspecification level (e.g., 9.6% of misspecification in this study) had larger bias. As for the balanced misspecification (i.e., some 1s were misspecified by 0s and the same number of 0s by 1s), attribute estimated bias slightly increased in a negative way as the misspecification percent increased. This may be because the effects of over- and under- misspecifications balance each other. Because over-misspecification exerted larger impact than under-misspecification, the balanced misspecification of the Q-matrix yielded negative bias.

As for RMSE in attribute estimates, its magnitude dramatically increased as the misspecification percent in the Q-matrix increased. Compared to the condition of no misspecification, the over-, under-, and balanced-misspecifications had substantial increases in RMSE but there were no differences in term of RMSE's magnitude between the three types of misspecification. As far as attribute estimate correlation between estimated and true parameters is concerned, there was no impact of misspecification factors. In other words, regardless of different types or percentages of misspecification in the Q-matrix, the attribute estimates and the corresponding true parameters were highly consistent, which means that the relative difficulties (or sample estimates' order) for cognitive attributes were consistent with the true attribute difficulties (or population parameters' order).

Overall, this simulation study suggests that misspecification type and percent do have a considerable impact on the bias and RMSE of attribute estimates in the crossed random effects LLTM using the SAS GLIMMIX procedure, especially under the conditions of high percent misspecification and of over-misspecification. Fortunately, attribute correlation between the estimated and true parameters is not affected by misspecification type and percent. In sum, since the Q-matrix is an indispensable element in applying the LLTM with item random effects, specifying an appropriate Q-matrix, therefore, is a crucial task and must be completed with generous assistance from content and subject experts.

REFERENCES

Cao, C., Wang, Y., Chen, Y.-H., & Li, I. Y. (2014). *Parameter estimation of cognitive attributes using the crossed-random effects linear logistic test model with PROC GLIMMIX*. *SAS Global Forum 2014 Proceedings.* Cary, NC: SAS Institute Inc.

Cassuto, N. (1996). *The performance of the linear logistic test model under different testing conditions*. Unpublished doctoral dissertation, University of Minnesota: Twin City, Minnesota.

Chen, Y.-H., MacDonald, G., & Leu, Y.-C. (2011). Validating cognitive sources of mathematics item difficulty: Application of the LLTM to fraction conceptual items. *The International Journal of Educational and Psychological Assessment, 7,* 74-93.

De Boeck, P. (2008). Random item IRT models. *Pschometrika, 73,* 533-559.

De Boeck, P. & Wilson, M. (2004). A framework for item response models. In P. De Boeck, & M. Wilson, (eds.), *Explanatory item response models: A generalized linear and nonlinear approach*, pp1-41. New York: Springer.

Embretson, S.E., & Daniel, R.C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science Quarterly, 50*, 328–344.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37,* 359–374.

Green, K. & Smith, R. M. (1987). A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics, 12,* 369-381.

Hartig, J., Frey, A., Nold, G., & Klieme, E. (2012). An application of explanatory item response modeling for model-based proficiency scaling. *Educational and Psychological Measurement, 72,* 665-686.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189-212). New York, NY: Springer.

Janssen, R., Tuerlinckx, F., Meulders, M., & DeBoeck, P. (2003). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics, 25,* 285-306.

Li, I. Y., Chen. Y.-H., & Kromrey, J. D. (2013). Evaluating the performance of the SAS® GLIMMIX procedure for dichotomous Rasch model: a simulation study. *Southeast SAS Users Group 2013 Proceedings.* Cary, NC: SAS Institute Inc.

MacDonald, G. (2014). *The performance of the linear logistic test model when the Q-matrix is misspecified: A simulation study.* Unpublished doctoral dissertation, University of South Florida: Tampa, Florida.

Rijmen, F. & De Boeck, P. (2002). The random weights linear logistic test model. *Applied Psychological Measurement, 26*, 271-285.

SAS Institute Inc. (2011). *SAS/STAT® 9.3 User's Guide: The GLIMMIX Procedure (Chapter 40)*. Cary, NC: SAS Institute Inc.

Wang, W.-C. & Jin, K.-Y. (2010). Multilevel, two-parameter, and random-weights generalizations of a model with internal restrictions on item difficulty. *Applied Psychological Measurement, 34,* 46-65.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Yi-Hsin Chen
Enterprise: University of South Florida
Address: 4202 E. Fowler Avenue, EDU105
City, State ZIP: Tampa, FL, 33620
Work Phone: (813) 974-4964
Fax: (813) 974-4495
E-mail: ychen5@usf.edu