Paper 83-14

Modeling Cognitive Processes of Learning with SAS® Procedures

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

ABSTRACT

Traditionally, the principal goal of educational assessments is to evaluate students' academic achievement or proficiency in comparison to their peers or against promulgated standards. Both classical test theory (CTT) and item response theory (IRT) modeling frameworks provide measurement in the form of a summative estimate of the outcome variable. In recent years, understanding and exploring the complex cognitive processes that contribute to the learning outcome received growing interest in the field of psychometrics, where various item response modeling approaches have been devised to explain the relationship between the outcome and its componential attributes. Such approaches include the linear logistic test model (LLTM), the crossed random-effects linear logistic test model (CRELLTM), and the multiple regression method (MR). This paper introduces these statistical models and illustrates their implementation with the GLM, NLMIXED, and GLIMMIX procedures over an empirical data set. A simulation study was conducted to examine the performance of these approaches.

INTRODUCTION

Faced with the prevalence of standardized tests which are either legislature-mandated (such as the No Child Left Behind Act of 2001) or commercially available, educators demand that the assessments their student are required to participate provide instructionally relevant results to have their worth maximized in helping with teaching and learning in return for their expensive costs. Advancement in cognition in educational psychology has increased our understanding of the mechanisms of human learning and problem solving, including cognitive domains, schemas, strategies, sub-operations/tasks, production rules, etc. As the traditional CTT and IRT frameworks are not suitable, the challenge to evaluate these unobservable processes has given rise to a series of componential models that are having a huge impact on test construction and analysis procedures. Before presenting these models, it is necessary to introduce briefly an indispensable modeling element.

Q-Matrix

The Q-matrix is an *I* by *K* incidence matrix mapping the *K* attributes to the *I* items. That is, each item is described by the cognitive processes upon which making a correct response depends. Typically, the columns of the Q-matrix are the attributes and the rows represent all the items. Also typically, an entry of "1" in the matrix indicates that attribute is needed to answer correctly the item on the row and an entry of "0" otherwise. In this case, these entries are called "indicators" (of occurrence). With some modeling approaches, there is dependency between the attributes or non-binary entries in the matrix as a fixed weighting system. The criticality of the Q-matrix lies in the fact that it paints the cognitive blueprint for test construction and componential models and must be specified before analysis. An example of the Q-matrix is given in Table 1.

The Linear Logistic Test Model

One of the early psychometric models incorporating cognitive processes is the linear logistic test model (LLTM) introduced by Fischer (1973), who discovered that an examinee must carry out eight basic cognitive operations in order to answer differentiating calculus questions. Presumably, the item difficulty parameter could be redefined in an additive function of the parameters of the eight cognitive processes plus person ability. In other words, the probability of answering a question correctly could be predicted using the LLTM by including person parameter and the item's cognitive properties. In traditional IRT framework, the LLTM is often considered a member of the Rasch families (Embretson & Reise, 2000). The fundamental formulation of the Rasch model for binary data can be expressed as

$$P\left(X_{ij}=1\middle|\theta_j,\beta_i\right)=\frac{\exp\left(\theta_j-\beta_i\right)}{\left[1+\exp\left(\theta_j-\beta_i\right)\right]},$$

where $P(X_{ij}=1)$ represents the probability of a correct response to item *i* for person *j*, θ_j is the ability level of person *j*, and β_i is the item difficulty of item *i*. When a linear summation of cognitive attributes replaces item difficulty, β_i, the Rasch model is extended into the LLTM, whose equation is given as:

$$P\left(X_{ij}=1\middle|\theta_j,q_{ik},\eta_k\right)=\frac{\exp\left(\theta_j-\sum_{k=1}^{K}q_{ik}\eta_k\right)}{\left[1+\exp\left(\theta_j-\sum_{k=1}^{K}q_{ik}\eta_k\right)\right]},$$

That is, in the LLTM the item difficulty parameters that are replaced by cognitive attributes can be conceptualized as

$$\beta_i = \sum_{k=1}^{K} q_{ik}\eta_k \, ,$$

where η_k is the process parameters of the model, whose values represent the contribution of cognitive attribute $k = 1, \ldots, K$ to the difficulty of item i; and q_{ik} is an entry in the pre-specified Q-matrix. In the framework of explanatory item response models (De Boeck & Wilson, 2004), the LLTM is considered an item explanatory model since item properties reflecting cognitive attributes are used to explain item responses. The LLTM can be implemented by the NLMIXED and GLIMMIX procedures as illustrated later.

The Crossed Random-Effects LLTM

The fundamental assumption of the LLTM is that the variation of the item difficulty parameter is accounted for completely by the process parameters behind an item, thus leaving no room for error in the model. Such an unrealistic relationship between the item parameter and its components is considered a limitation with the LLTM (Hartig, Frey, Nold, & Klieme, 2012). Studies (Janssen and De Boeck, 2002; Rijmen & de Boeck, 2002) found that the model-data fit of the LLTM was less than satisfactory even when the Q-matrix is well designed and most of the variance in item difficulty parameters is explained by parameters of cognitive attributes. By adding an error term on item difficulty (i.e., $\hat{\beta}_i = \beta_i + \varepsilon_i$), the crossed random-effects LLTM (CRELLTM) proposed by Janssen and colleagues (2004) resolves this limitation.

With the CRELLTM, the item difficulty parameters as predicted by parameters of cognitive processes can be conceptualized as

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

where ε_i reflects the portion of difficulty parameter for item i unaccounted for by cognitive attributes. The accuracy and precision of parameter estimation for cognitive attributes under the CRELLTM was investigated as implemented with the GLIMMIX procedure (Cao, Wang, Chen, & Li, 2013). With generalized linear mixed models, this error term is captured by treating item parameters as a random effect.

The Regression Approach

Given that item difficulty is defined in the LLTM as a function of all parameters of cognitive attributes, a crude method to obtain these parameter estimates is through a multiple regression. First, item difficulty parameters are estimated with the GLIMMIX procedure (Chen, Li, & Kromrey, 2013) by fitting the dichotomous Rasch model. Next, the item parameter estimates are regressed on all attributes represented by their occurrence in the Q-matrix. That is, for the difficulty estimate of a certain item, its predictors include only those attributes contributing to it and their values are all one.

AN EXAMPLE

560 fourth graders from a Midwestern state sat in an administration of a widely used standardized test designed to provide instructional support for schools. Only the 19 reading comprehension questions from the exam were used. The three processes of comprehension assessed by the exam include retrieval and understanding, inference and interpretation, and evaluation and synthesis. A group of subject matter experts studied the items and mapped them to the three processes to create a plausible Q-matrix, which is shown in Table 1 below. The response data were transformed into the “long format”, where all responses filled one column ordered by students and items. An excerpt of the data in this format is given in Table 1, which shows only the segment for Student 1. Column 3 contains his/her responses (“scores”) to all 19 items. For example, the response to item2 was correct but that to item 3 was incorrect. Column “c1” to “c3” represent the three processes of the Q-matrix, which is repeated for every segment of the data for all 560 students. Note that some items tap into more than one comprehension process (e.g., the third and eighth items both assessing retrieval and understanding as well as inference and interpretation). That is, students are expected by the model to implement the two processes in order to answer the question correctly.

Methods

As discussed earlier, the LLTM can be considered an item explanatory model and thus a case of generalized linear mixed models (GLMMs). Therefore both the NLMIXED and GLIMMIX procedures can be used to fit this model. The

NLMIXED procedure in SAS for the non-linear mixed modeling (NLMMs), which subsumes GLMMs, has been widely used to formulate various item response theory (IRT) models (e.g., De Boeck & Wilson, 2004). This procedure becomes more popular for researchers and practitioners because of the availability of SAS and modeling flexibility due to its capability to include item discrimination parameter. The LLTM code as implemented with NLMIXED is follows.

```
proc nlmixed data=all qpoints=20 tech=newrap;
  parms b1-b3=0 sd0=1;
  beta= b1*c1 + b2*c2 + b3*c3;
  ex=exp(theta-beta);
  p=ex/(1+ex);
  model response ~ binary(p);
  Random theta ~ normal(0,sd0**2) subject=student OUT = theta_hat;
  Estimate 'sd0**2' sd0**2;
  ods output Parameters = Parmsnl ParameterEstimates = Parmnl AdditionalEstimates =
AddEstnl FitStatistics = Fitnl ;
 run;
```

Table 1 Excerpt of the Response Data

Student	Item	Response	c1	c2	c3
1	i1	1	0	0	1
1	i2	1	0	0	1
1	i3	0	1	1	0
1	i4	1	1	0	0
1	i5	1	0	1	0
1	i6	1	0	1	0
1	i7	1	0	1	0
1	i8	1	1	1	0
1	i9	0	1	0	0
1	i10	1	1	0	0
1	i11	1	1	1	1
1	i12	0	0	1	1
1	i13	1	1	0	0
1	i14	1	1	0	0
1	i15	1	0	1	0
1	i16	0	0	1	0
1	i17	0	1	0	0
1	i18	0	0	1	0
1	i19	1	0	1	1

Data set “all” is in the long format. The Newton-Raphson (newrap) technique is used for optimization, which requires repeated computation of the function value (optimization criterion) and gradient vector (first-order partial derivatives). Other optimization techniques are available with NLMIXED. The integration method here is the adaptive gaussian quadrature The process parameters b1 to b3 are given an initial value of to 0 and initial standard deviation of 1 and an estimate of their variance is requested . In this example, instead of the original conditional maximum likelihood method, a variant of the LLTM (De Boeck and Wilson, 2004) is applied that employs marginal maximum likelihood estimation, which is available with PROC NLMIXED. In model definition, the item parameter beta equals to the sum of the three processes multiplied by their coefficients (b1 to b3), the parameters of interest. Note that the student ability, theta, assumes a normal distribution with a mean of 0 and a standard deviation of 1. Finally, output is saved in specified tables via the ODS. Parameter estimates of b1 to b3 are given below. Their values reflect the effect of every process on item difficulty. The negative values of the first two processes are expected since they are the “easier” ones.

Table 2 LLTM Estimates Using NLMIXED

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
b1	-0.3740	0.03991	559	-9.37	<.0001	0.05	-0.4524	-0.2956	-4.78E-7
b2	-0.5124	0.03846	559	-13.32	<.0001	0.05	-0.5879	-0.4369	-6.07E-7
b3	0.2193	0.04756	559	4.61	<.0001	0.05	0.1259	0.3127	-1.8E-7
sd0	0.8473	0.03848	559	22.02	<.0001	0.05	0.7717	0.9229	-4.4E-6

The GLIMMIX procedure is a relatively new package in SAS and has been shown as a good tool for psychometric analysis (Li, Chen, & Kromrey, 2013). The SAS PROC GLIMMIX performs estimation and statistical inference for GLMMs that extend the class of generalized linear models (GLMs) by incorporating normally distributed random effects. The GLIMMIX procedure has been applied to implement a number of item response models and is gaining increasing popularity in other fields due to its powerful functionalities and the wider use of GLMMs.

The GLIMMIX and NLMIXED procedures conduct parameter estimation as an optimization problem and approximate the marginal likelihood as the solution. NLMIXED uses an integral approximation through Gaussian quadrature. GLIMMIX by default fits GLMMs based on linearizations with a technique known as restricted pseudo-likelihood (RPL) but it is also capable of estimation methods based on integral approximation, a Laplace approximation of the marginal log likelihood and adaptive Gauss-Hermite quadrature. The code to implement the LLTM with GLIMMIX is here.

```
proc glimmix data = all method=quad;
 class item student;
 model response (event='1')= c1 c2 c3/ solution noint link=logit dist = binary ;
 random int / subject = student solution;
 ods output FitStatistics=Fitlltm ParameterEstimates = Parmlltm SolutionR=Rlltm;
run;
```

Maximum likelihood estimation and quadrature method of likelihood approximation are used. The MODEL statement lists the three attribute parameters as the fixed effects and predictors for a correct response. Their estimates are given in Table 3, where the direction of the effects will have to be reversed. The RANDOM statement identifies students (theta) as random effects in the mixed model and requests its estimate by specifying the SOLUTION option.

Table 3 LLTM Estimates Using GLIMMIX

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
c1	0.3742	0.03988	10077	9.38	<.0001
c2	0.5125	0.03843	10077	13.34	<.0001
c3	-0.2192	0.04754	10077	-4.61	<.0001

The CRELLTM adds an error term in the regression function of item difficulty parameter on attribute parameters. GLIMMIX realizes that by also identifying items (beta) as random effects as shown in the following code.

```
proc glimmix data = all method=laplace;
 class item student;
 model response (event='1') = c1 c2 c3/ solution noint link=logit dist = binary  ;
 random int / subject = student solution;
 random int / subject = item solution;
 ods output FitStatistics= Fitcrelltm ParameterEstimates = Parmcrelltm
SolutionR=Rcrelltm;
run;
```

Here the most obvious difference from the LLTM code is the addition of another RANDOM statement. As both items and students in the CLASS statement are treated as random effects, this belongs with the category of crossed models, which is in fact a version of explanatory IRT models that account for the variation of both item difficulty and person ability. However, the NLMIXED procedure cannot be used to estimate random effects for person ability and item difficulty simultaneously with quadrature approximation method (De Boeck & Wilson, 2004). Similarly, this method when utilized by GLIMMIX does not work either. Therefore the approximation method used here is Laplace with optimization technique being dual quasi-newton.

The attribute parameter estimates as fixed effects from the CRELLTM are given in Table 4. Note the similar but not close values to those in Table 3 and the need for reversing the direction of the fixed effects. Table 5 lists the model fit statistics output by GLIMMIX for the LLTM (on the left) and the CRELLTM (on the right). The CRELLTM had better fit.

Table 4 CRELLTM Estimates

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
c1	0.4778	0.2482	10062	1.93	0.0542
c2	0.7067	0.2430	10062	2.91	0.0036
c3	-0.2262	0.3453	10062	-0.66	0.5123

The final approach to obtain process parameter estimates is through a regression analysis after item difficulties have been estimated. When items are considered nested within persons, the dichotomous Rasch model can be seen as a member of the GLMMs, enabling it to be fitted with GLIMMIX. The SAS code to derive estimates both item and person (i.e. students) parameters is as follows.

```
proc glimmix data = all method=quad;
 class item student;
 model response (event='1') = item student / solution noint link=logit dist = binary;
 lsmeans item / cl ilink;
 lsmeans student / cl ilink;
 ods output lsmeans=lsmeans_item_effects FitStatistics= FitRasch  ;
run;
```

The marginal log likelihood of the Rasch model is approximated using maximum likelihood estimation with Newton-Raphson being the optimization technique. The direction of item effects output by the LSMEANS statement needs to be reversed for them to be interpreted as item difficulties. Table 6 presents the estimates of item difficulties from the four approaches. Note that those values from the two LLTM approaches are calculated on the basis of previous formulae whereas the CRELLTM and the Rasch estimates were provided by the GLIMMIX procedure.

Table 5 Fit Statistics from GLIMMIX for LLTM and CRELLTM

Fit Statistics	
-2 Log Likelihood	13361.53
AIC (smaller is better)	13369.53
AICC (smaller is better)	13369.54
BIC (smaller is better)	13386.84
CAIC (smaller is better)	13390.84
HQIC (smaller is better)	13376.29

Fit Statistics	
-2 Log Likelihood	12560.96
AIC (smaller is better)	12570.96
AICC (smaller is better)	12570.97
BIC (smaller is better)	12560.96
CAIC (smaller is better)	12565.96
HQIC (smaller is better)	12560.96

Table 6 Item Difficulty Estimates from the Four Approaches

Item	NL LLTM	LLTM	CRELLTM	Rasch
i1	0.2193	0.2192	-1.0728	-1.1253
i2	0.2193	0.2192	-0.6437	-0.6425
i3	-0.8864	-0.8867	-0.3433	-1.8548
i4	-0.3740	-0.3742	-0.1274	-0.8400
i5	-0.5124	-0.5125	-1.5522	-2.6958
i6	-0.5124	-0.5125	-0.5974	-1.6173
i7	-0.5124	-0.5125	-0.5974	-1.6173
i8	-0.8864	-0.8867	-0.2966	-1.8028
i9	-0.3740	-0.3742	0.3002	-0.3588
i10	-0.8864	-0.8867	0.4361	-0.2052
i11	-0.6671	-0.6675	0.3361	-0.8496
i12	-0.2931	-0.2933	0.3107	-0.3498
i13	-0.3740	-0.3742	-0.4275	-1.1765
i14	-0.3740	-0.3742	-0.0514	-0.7546
i15	-0.5124	-0.5125	0.5092	-0.3770
i16	-0.5124	-0.5125	0.5413	-0.3407
i17	-0.3740	-0.3742	0.4361	-0.2052
i18	-0.5124	-0.5125	0.8927	0.0575
i19	-0.2931	-0.2933	1.1391	0.5943

As expected, the two LLTM approaches supplied very close item difficulty estimates. However, they are far from the estimates from the other two approaches, confirming the less than realistic assumption of the LLTM with regards to the item parameter. The estimates based on the CRELLTM and on the Rasch model are similar but not close, suggestive of the fact that the CRELLTM includes the effects of cognitive processes in its analysis in addition to student scores.

The estimates of person ability for the 560 students as gauged by the reading comprehension questions are compared between the approaches in the graph below. Note that these ability estimates are in ascending order.

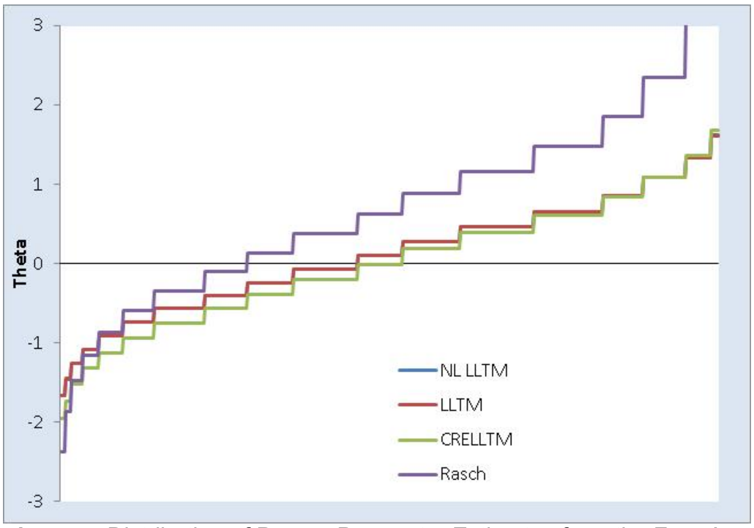

Figure 1: Distribution of Person Parameter Estimates from the Four Approaches

It is clear that the LLTM estimates using NLMIXED and GLIMMIX are so close that their lines overlap. These ability estimates are very close to the CRELLTM ones, especially towards the top end of the spectrum. On the other hand, the Rasch ability values are distant from the other three sets and are increasingly higher as the ability level goes up. The person parameter is fixed for the Rasch analysis but treated as random effects for the other three approaches defined as $\theta_j \sim N(0, \sigma_\theta^2)$.

After, the effects from the cognitive attributes can be estimated by regressing item difficulty estimates on the indicators in the Q-matrix. That is, the values of the three predictors are either 1 or 0. This linear regression can be conducted using GLIMMIX as follows.

```
proc glimmix data = mvreg ;
 model difficulty = c1 - c3 / solution noint;
 ods output parameterestimates = parmreg fitstatistics = fitreg ;
run;
```

Apparently this is a crude method since it excludes information on individual differences in ability. The R-square value is only .56. Table 7 below lists the estimates for the three process parameters. Clearly, the values are not close to those obtained with the other approaches.

Table 7 Estimates Based on Regression

Parameter Estimates					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
c1	-0.6255	0.2928	16	-2.14	0.0484
c2	-0.8618	0.2863	16	-3.01	0.0083
c3	0.1676	0.4095	16	0.41	0.6878
Scale	0.7006	0.2477	.	.	.

SIMULATION STUDY

Which is the preferred approach to model cognitive processes? To answer this question, a comprehensive simulation study was conducted under a number of conditions. Data were simulated using a random number generator, RANNOR, in SAS/IML package. For each condition, 1000 replications were generated. The manipulated sample sizes were 25, 50, 100, 250, and 500. The population distribution shapes were set up as follows: normal (skewness = 0, kurtosis= 0), negatively-skewed distribution (skewness = -2, Kurtosis = 6), slightly negatively-skewed (skewness = -1, Kurtosis = 3), slightly positively-skewed (skewness = 1, kurtosis = 3), positively-skewed (skewness = 2, kurtosis = 6), highly leptokurtic (skewness = 0, kurtosis = 25), and slight platykurtic (skewness = 0, kurtosis = -1). The numbers of items was fixed at 21. It was decided to select eight cognitive attributes. The forms of the Q-matrix density was sparse or dense, the details of which are not discussed.

Two decision criteria used to assess the estimation performance of these approaches are estimation bias and root mean square error (RMSE). Estimation bias is computed as the average difference between the estimated and true parameters:

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

where η is the attribute parameter. 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. It is defined as

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

In addition, factorial ANOVA analysis with the generalized eta-squared effect size was conducted to examine what manipulated factors affected bias and RMSE significantly.

Results

Overall, the CRELLTM approach performed better in terms of parameter estimation of cognitive attributes. The overall distribution of estimation bias for all four approaches is presented as the box and whisker plot in Figure 2. Tthe box plots of the two LLTM methods are almost identical. The average bias for CRELLTM method is smaller (-0.0321) than those of the LLTM methods (-.0691 and -.0687). The variance of the estimation bias for CRELLTM is also lower (SD = .1665) than those of LLTM (.1788 and .1791). In comparison, estimation bias and its variance associated with the regression approach are far greater (Mean = .0118 and SD = .4910).

The factorial ANOVA analysis provided generalized eta-squared effect size for all four methods, which are summarized in Table 7. The largest effect for CRELLTM and the two LLTM methods came from the interaction between sample size and population distribution shape (underlined values), which are significantly higher than the rest of the effects. As for the regression approach, significant effects were found from sample size, population distribution shape, and their interaction.

Table 8 Eta Squared Effect Size from Attribute Bias

Source	CRELLTM	LLTM-GLIMMIX	LLTM-NLMIXED	Regression
N_PERSONS*SHAPE	<u>0.3607</u>	<u>0.3765</u>	<u>0.3767</u>	<u>0.2344</u>
SHAPE*QF	0.0747	0.0882	0.0881	0.0119
N_PERSONS	0.0670	0.0603	0.0601	<u>0.3594</u>
N_PERSONS*QF	0.0523	0.0395	0.0395	0.0410
SHAPE	0.0399	0.0448	0.0452	<u>0.2614</u>
QF	0.0234	0.0131	0.0130	0.0384

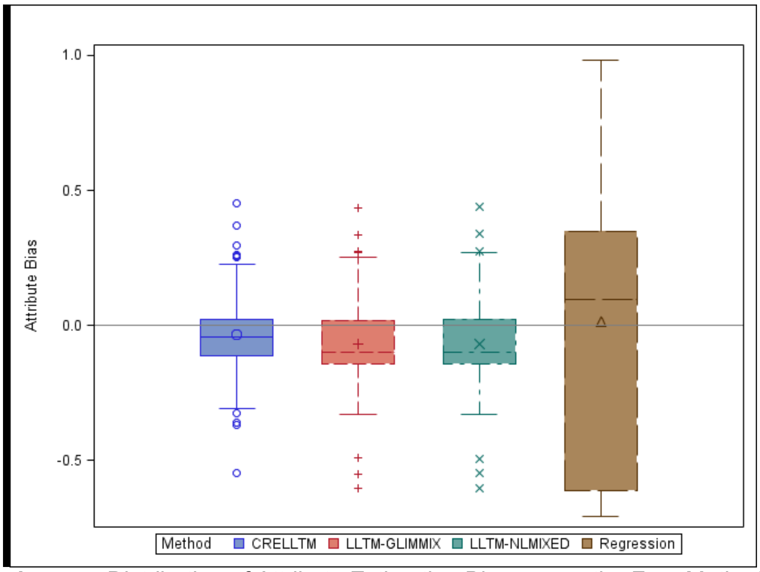

Figure 2: Distribution of Attribute Estimation Bias across the Four Methods

The average RMSE of attributes estimation for all four approaches is quite large (.6135 for CRELLTM, .6449 for LLTM-GLIMMIX, and .6456 for LLTM-NLMIXED). This appears to contradict the findings of relatively small bias from the three non-regression approaches as discussed in the previous section. Figure 3 displays the distributions of RMSE. Again, CRELLTM exhibited smaller and less varying RMSE than the other methods whereas the two LLTM approaches had almost identical distributions. The RMSE for the regression method is far greater than others.

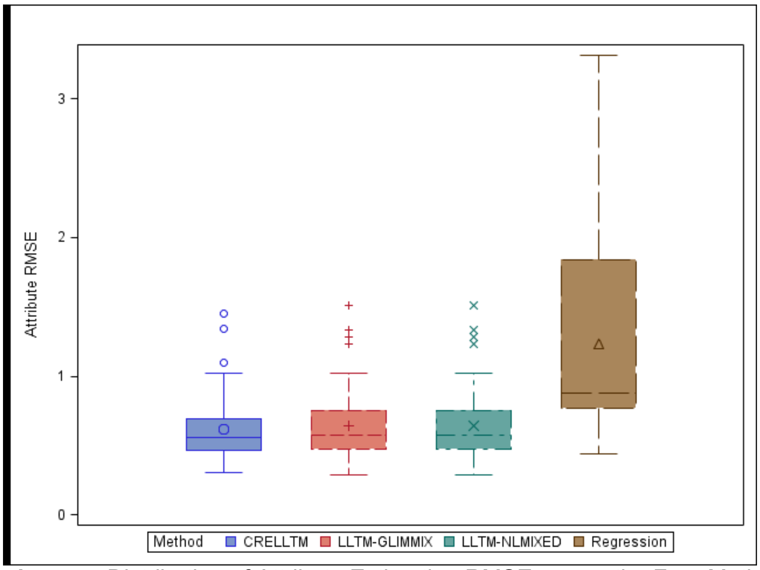

Figure 3: Distribution of Attribute Estimation RMSE across the Four Methods

The results of factorial ANOVA analysis with generalized eta-squared effect sizes indicates that the main effect of the interaction between sample size and population distribution shape has a considerable effect on the RMSE of the attributes estimates from the three non-regression methods. All three η^2 values are at a great level (> .3300). In addition, the effect size of interaction between population distribution shape and Q-matrix density is also large for the CRELLTM approach. On the other hand, the effect size of Q-matrix density is significant for the two LLTM methods as exhibited in Figure 4. In Table 2, all these η^2 are underlined.

Table 9 Eta Squared Effect Size from Attribute RSME

Source	CRELLTM	LLTM-GLIMMIX	LLTM-NLMIXED	Regression
N_PERSONS*SHAPE	<u>0.3355</u>	<u>0.3538</u>	<u>0.3534</u>	0.0518
SHAPE*QF	<u>0.1119</u>	0.0754	0.0755	0.0405
N_PERSONS	0.0909	0.0235	0.0237	
QF	0.0493	<u>0.1475</u>	<u>0.1486</u>	
SHAPE	0.0412	0.0467	0.0464	0.0114
N_PERSONS*QF	0.0338	0.0316	0.0315	

CONCLUSION

When componential modeling, including study of cognitive processes, receives growing attention in social and behavioral sciences, it is exciting that SAS has acquired powerful functionalities to implement analyses of this kind with procedures of NLMIXED and GLIMMIX. This paper has illustrated how to fit several cognitive process models using these procedures. The LLTM as implemented with NLMIXED and GLIMMIX gave estimates that were almost identical. Since it is less than reasonable to assume that the attribute parameters account for the entire variation of the item parameter, the CRELLTM is a better modeling approach by adding an error term. GLIMMIX is the only procedure capable of fitting it with the additional error term by allowing items to be random effects in the mixed model. On the other hand, the two-step regression approach using GLIMMIX is the least favored given that each step produces error in estimation. Li et al. (2013) found that GLIMMIX running Rasch analysis can run into large errors with small sample sizes and short tests. Moreover, the step of regressing estimated item parameters on the Q-matrix is problematic since it ignores personal differences in the data.

The evidence from the simulation study supports the reasoning. The CRELLTM performed the best in terms of estimating the effects from the cognitive attributes. The average estimation bias and RMSE across all conditions is negligible. But it should be cautioned that under some conditions considerable bias and large deviances occurred. Factorial ANOVA analysis with generalized eta-square effect sizes reveals that the most influential main effect (all effect sizes > .3) comes from the interaction between sample size and population distribution shape consistently for the three non-regression approaches. Another notable effect is due to density of the Q-matrix.

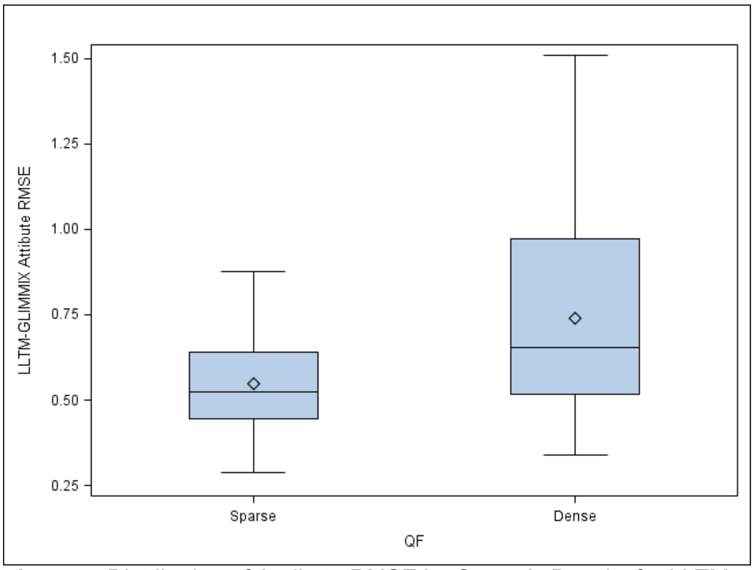

Figure 4: Distribution of Attribute RMSE by Q-matrix Density for LLTM

REFERENCES

Cao, C., Wang, Y., Chen, Y. H., & Li, I. (2013). Parameter Estimation of Cognitive Attributes using the Crossed Random-Effects Linear Logistic Test Model with PROC GLIMMIX. *Proceedings of the Annual SAS Users Global Forum.* Cary, NC: SAS Institute Inc

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.

DeBoeck, P., & Wilson, M. (2004). *Statistics for Social Science and Public Policy: Explanatory Item Response Models*. Springer-Verlag: New York, New York.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Psychology Press.

Fischer, G. H. (1973). 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., & De Boeck, P. (2000). *A random effects version of the linear logistic test model.* Manuscript submitted for publication.

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. *Proceedings of the Annual Southeast SAS Users Group Conference.* Cary, NC: SAS Institute Inc.

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*. 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: Isaac Li
Enterprise: University of South Florida

Address: 4202 E. Fowler Avenue, EDU105
City, State ZIP: Tampa, FL, 33620
E-mail: liy1@mail.usf.edu