

# Applying Linear Mixed Effects Models with Crossed Random Effects to Psycholinguistic Data: Multilevel Specification and Model Selection

Hsiu-Ting Yu <sup>a</sup>

<sup>a</sup> McGill University

**Abstract** ■ Applying linear mixed effects regression (LMER) models to psycholinguistic data was made popular by Baayen, Davidson, and Bates (2008). However, applied researchers sometimes encounter model specification difficulties when using such models. This article presents a multilevel specification of LMERs customized for typical psycholinguistic studies. The proposed LMER specifications with crossed random effects allow different combinations of random intercept effects or random slope effects to be specified directly for subject and item covariates. As a result, this approach allows researchers to describe, specify, and interpret a wide range of effects in an LMER more easily. Next, the syntax and steps involved in using the PROC MIXED procedure in SAS to fit the discussed models are illustrated. Thirdly, various issues relating to model selection, specifically for the random component of LMER models with crossed random effects, are discussed. Finally, this article concludes with remarks about model specification and selection of the random structure in the context of analyzing psycholinguistic data using LMERs specifically. This paper provides readers conducting psycholinguistic research with a complete tutorial on how to select, apply, and interpret the multilevel specification of LMERs.

**Keywords** ■ Model selection, linear mixed effects model, random effects, psycholinguistics

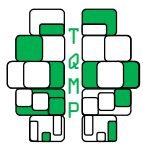
 [ht.yu@mcgill.ca](mailto:ht.yu@mcgill.ca)

## Introduction

In a typical psycholinguistic study, the experimenter presents multiple test items to multiple subjects, and the subjects respond to each item. The classical analysis of such data uses two analysis of variance (ANOVA) tests:  $F_1$  analysis (by subject) and  $F_2$  analysis (by item) (Clark, 1973).  $F_1$  and  $F_2$  analyses investigate the condition means obtained for each subject and each item, respectively, and the  $F_1 \times F_2$  criterion is achieved when both analyses are significant. This standard approach implicitly treats the materials used in an experiment as fixed factors, but Clark (1973) coined the term "language-as-fixed-effect fallacy" to argue that random effects may be more appropriate for analyzing linguistic data. The quasi-F ratio ( $F'$ ) is a random effects model that takes into account both item and subject variability, but this ratio cannot be computed when the data are unbalanced or when responses are missing for certain item/subject combinations. Therefore, the  $\min F'$  (Clark, 1973) was proposed to be the minimum bound of the  $F'$  and was used as a criterion in the analysis.

The application of linear mixed-effects regression (LMER) models to psycholinguistic data was recently made popular by Baayen, Davidson, and Bates (2008).

They offer a general introduction to the modeling framework and discuss the advantages LMER has over traditional methods. They describe how models can be fitted using the *lmer* package in R (R Development Core Team, 2008). Since the time that this introductory paper and free software was made available, LMER has rapidly become a preferred analysis method in psycholinguistics and related fields (e.g., Kuperman & Van Dyke, 2011; Moreton, 2012; Staub, 2009). The popularity of the method is evidenced by the fact that since 2007, the Baayen and colleagues' article (2008) is the most cited article in the *Journal of Memory and Language* (JML), thus demonstrating the method's popularity. In addition, from November 2008 (vol. 59, issue 4) through November 2012 (vol. 67, issue 4), 67 (22%) of the articles in JML used LMER or related techniques. A number of studies have demonstrated the various advantages mixed effects models have over traditional random-effects ANOVAs (e.g., Baayen et al., 2008; Quené & van den Bergh, 2008). The key in the LMER is to have fixed and random effects in the model to account for the variation observed in the data. The different types of effects in LMER serve the context of psycholinguistic studies well and are discussed next.



Experimental factors can be viewed as arising from effects that are either fixed or random. Different treatment manipulations are usually treated as different levels of a fixed effect. When an effect is treated as fixed, the inferential results are valid for the particular levels selected in the experiment. As information is limited to a particular sample, the error is usually underestimated when generalizing to another sample (e.g., the estimated confidence intervals are too narrow). Hence, assuming an effect is fixed when it is actually random produces overly optimistic estimates of the reliability of the effect in a new sample, resulting in inflated Type 1 error rates. Erroneously treating either subject or item effects as fixed in a recognition memory paradigm can cause surprisingly large Type 1 error rates (Baayen et al., 2008; Rouder, 2005).

In the context of analyzing psycholinguistic data using LMERs, random effects can be imposed on subjects and/or items. Including *random subject effects* is quite common and easy for psychologists because individual differences are the fundamental assumption in psychological studies. Situating this concept in a statistical context, random subject effects can be conceptualized as subjects that are sampled with replacements from a parametric distribution. For example, the gender (male, female) can be included as a subject covariate, and gender differences can be considered a systematic random subject effect. On the other hand, an example of *random item effects* can be found in word recognition studies. When a set of words is presented to subjects, these words may vary in some lexical dimensions of interest (e.g., phonological neighborhood density). As an example, the word frequency (high, low) can be treated as an item covariate, and the familiarity of the word stimuli can lead to systematic random item effects.

Variations due to individual differences are usually treated as random effects. Random effects differ from fixed effects in several ways. First, instead of point estimates for parameters, the distribution (usually the variance) of the effect is estimated. Second, instead of assuming effects at specific levels, the random effects "explain" variations in the data. These two characteristics lead to the possible usages of random effects. If the degree of variation among the subjects or among the items is the research question, then estimating that variation and its structure is the primary task. On the other hand, the random effects can be treated as nuisance variables if the random component is used to account for as much variation

(among items and/or subjects) as possible in the data.

### The Specification of LMER in Baayen et al. (2008)

The specification of LMER promoted by Baayen et al. (2008) has the form of a regression equation, and the parameters and covariates are in matrix form. Specifically, it has the following form:

$$y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{S}_i\mathbf{s}_i + \mathbf{W}_j\mathbf{w}_j + \varepsilon_{ij}, \quad (1)$$

where the vector  $\mathbf{y}_{ij}$  represents the responses of subject  $i$  to item  $j$ , and  $\mathbf{X}_{ij}$  denotes the design matrix consisting of ones in the first column and covariates in the other columns.  $\mathbf{S}_i$  and  $\mathbf{W}_j$  denote *subject* and *item* matrices, which are identical to the specifications of  $\mathbf{X}_{ij}$  and  $\boldsymbol{\beta}$  is the vector of the coefficients associated with  $\mathbf{X}_{ij}$ .  $\mathbf{s}_i$  and  $\mathbf{w}_j$  are subject and item random effects assumed to be normally distributed with mean 0 and variance  $\sigma_s^2$  and  $\sigma_w^2$ , respectively, and  $\varepsilon_{ij}$  is the normally distributed residual term with mean 0 and variance  $\sigma^2$ .

The specification in Equation (1) is not particularly intuitive or easy for applied researchers for the following reasons: (1) the subject and item covariates are both included in the design matrix ( $\mathbf{X}_{ij}$ ), although their effects are preferred to be discussed separately in practice; (2) the random intercept and random slope are not clearly represented in terms of whether they are associated with subject or item covariates; and (3) the vector of ones in the first column of  $\mathbf{X}_{ij}$ ,  $\mathbf{S}_i$  and  $\mathbf{W}_j$  may not be familiar to applied researchers. Therefore, an alternative specification that is rooted in multilevel modeling is proposed and illustrated in this paper.

The multilevel modeling techniques applied to item analysis in psycholinguistic research have been discussed in Locker, Hoffman, and Bovaird (2007). However, the specification proposed here uses a two-layer representation to describe different effects. The first layer describes the relationship between the outcome and the covariates of interest. The second layer describes how the parameters of the first layer relate to the fixed and/or random components. This two-layer representation offers a simple and explicit presentation of the various effects researchers may wish to test.

### Multilevel Specifications of an LMER model

The LMER offers a better approach to address the three issues of using the traditional univariate ANOVA approach. In this section, a multilevel specification of

**Table 1** ■ Example of Long Format Data Structure

| SubID | ItemID | $x_{i1}$ | $x_{j1}$ | $y_{ij}$ |
|-------|--------|----------|----------|----------|
| 1     | 1      | 1        | -0.5     | 123      |
| 1     | 2      | 0        | -0.5     | 145      |
| 1     | 3      | 0        | -0.5     | 225      |
| ⋮     | ⋮      | ⋮        | ⋮        | ⋮        |
| 1     | 10     | 1        | -0.5     | 165      |
| ⋮     | ⋮      | ⋮        | ⋮        | ⋮        |
| ⋮     | ⋮      | ⋮        | ⋮        | ⋮        |
| ⋮     | ⋮      | ⋮        | ⋮        | ⋮        |
| 10    | 1      | 1        | 0.5      | 221      |
| 10    | 2      | 0        | 0.5      | 167      |
| 10    | 3      | 0        | 0.5      | 196      |
| ⋮     | ⋮      | ⋮        | ⋮        | ⋮        |
| 10    | 10     | 1        | 0.5      | 246      |

LMER models is described. The SAS PROC MIXED syntax for fitting each of the discussed models is also included.

The multilevel framework consists of two layers. The first layer specifies a multiple linear regression model for the outcome variable where the subject and item covariates are represented as different terms. Specifically, the first layer is represented as

$$y_{ij} = \beta_0 + \beta_i x_{i1} + \beta_j x_{j1} + \varepsilon_{ij} \quad (2)$$

where  $\varepsilon_{ij}$  is the residuals, as specified earlier in Baayen et al. (2008),  $\beta_0$  is the intercept coefficient, and  $\beta_i$  and  $\beta_j$  are the slope coefficients for the subject covariate ( $x_{i1}$ ) and item covariate ( $x_{j1}$ ), respectively. For ease of illustration and interpretation, the illustration used here includes only one subject covariate and one item covariate<sup>1</sup>, and each has two levels. For example, two levels of the subject covariate ( $X_{i1}$ ) might be participant gender (male, female), and two levels of the item covariate ( $X_{j1}$ ) might be word frequency (high, low). Both covariates are coded using an effect coding scheme (e.g., -0.5 and 0.5). Centering at zero under the effect coding scheme allows easier and more intuitive interpretation of the effects, especially in interpreting

the average effects and the effects between levels.

In practice, studies have multiple subject and item covariates. The multiple covariates can be represented by adding additional terms to Equation (2). Specifically, the terms ( $X_{i1}, X_{i2}, \dots, X_{ip}$ ) and ( $X_{j1}, X_{j2}, \dots, X_{jq}$ ) can be added in the first layer, with the associated coefficients representing the  $p$  subject and  $q$  item covariates.

The second layer specifies the effects associated with the parameters in the first layer (i.e.,  $\beta_0, \beta_i$ , and  $\beta_j$ ). The simplest case, a model without random effects, is specified using these equations:

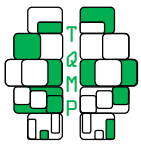
$$\begin{aligned} \beta_0 &= \gamma_0 \\ \beta_i &= \gamma_{i.} \\ \beta_j &= \gamma_{.j} \end{aligned} \quad (3)$$

The model in Equation (3) is usually referred to as the *baseline model* or *fixed effects only* model because no random effects are included. The intercept ( $\beta_0$ ) is the sum of the intercept for the subjects ( $\gamma_{0i}$ ) and the intercept for the items ( $\gamma_{0j}$ ). But these two coefficients cannot be estimated separately, so the  $\gamma_0$  is used in the equation to represent the additive effect of intercepts from subjects and items.  $\gamma_{i.}$  and  $\gamma_{.j}$  represent the mean effect of the subject covariate ( $X_{i1}$ ) and the item covariate ( $X_{j1}$ ) on the dependent variable  $y$ , respectively, after controlling for all other terms in the equation.

### Random Intercept Models

When there are systematic variations in data (e.g., due to participant gender or the experimental manipulations), random effects can be included to

<sup>1</sup> The terms “covariates,” “predictors,” and “factors” are used interchangeably in preferring the  $X$  in the model to represent the experimental design, subject characteristics, or item properties. In this paper, the term “covariate” is used.



capture the patterns of variation. Variation can be associated with the subject characteristics, the item characteristics, or both. Let  $S$  and  $I$  denote independent random variables with the same assumed distribution that capture subject- and item-associated variation. More specifically, the random intercept effects for subjects ( $S_{0i}$ ) and items ( $I_{0j}$ ) are assumed to be normally distributed with a mean of zero and a variance of  $U_{0i}^2$  and  $V_{0j}^2$ , respectively. They are also assumed to be independent from the residuals ( $\varepsilon_{ij}$ ).  $S_{0i}$  and  $I_{0j}$  are also called subject and item residuals because they represent the degree of deviation from the overall mean.

When we specify an intercept coefficient for subjects as  $\beta_0 = \gamma_0 + S_{0i}$ , we have a *random subject intercept model*. By including a subject coefficient, we allow the outcome variable to vary by subject. The interpretation of such a model is that when all other factors are kept constant, subject  $i$  can deviate from the mean of the outcome variable  $Y$  by  $s_{0i}$ , where  $s_{0i}$  is a realization of the random variable  $S_{0i}$ .

Using PROC MIXED to fit this model requires the data to be structured in long format. Specifically, each row contains information about the responses (e.g., outcome, covariates, subject ID, item ID, etc.) of a trial. In other words, the number of rows in the data file for each subject equals the number of trials that subject responded to. An example of the long format data structure is illustrated in Table 1.

The *random subject intercept model* described earlier can be represented in this two-layer specification:

$$\begin{aligned}y_{ij} &= \beta_0 + \beta_i x_{i1} + \beta_j x_{j1} + \varepsilon_{ij} \\ \beta_0 &= \gamma_0 + S_{0i} \\ \beta_i &= \gamma_{1.} \\ \beta_j &= \gamma_{.1}\end{aligned}\quad (4)$$

The *random intercept effects for subject* can be indicated by including the subject identification variable (SubID) in the random statement. Specifically, the PROC MIXED syntax for this model is:

```
PROC MIXED data=Data1 method=REML ic;
CLASS SubID ItemID;
MODEL y = X_i X_j /solution;
RANDOM intercept /subject=SubID;
RUN;
```

Like several other modeling procedures in SAS,

PROC MIXED includes a **CLASS** statement and a **MODEL** statement. The **solution** option is used to request the printing of the estimates for all the fixed effects in the model together with standard errors,  $t$ -statistics, and corresponding  $p$ -values for testing their significance. The **CLASS** statement names the classification variables to be used in the model. The **MODEL** statement is used to specify the fixed effects in the model. In this example, they are  $X_i$  and  $X_j$  subject and item covariates. The **RANDOM** statement is where users can specify the random effects they want to estimate. The random intercept effects for subjects ( $S_{0i}$ ) is specified in the **RANDOM** statement: **RANDOM** intercept /subject=SubID;. The **subject=** option in the **RANDOM** statement is used to indicate the clustering/grouping variables.

The same approach applies to the items; the model that includes random intercept effects for items,  $\beta_0 = \gamma_0 + I_{0j}$ , implies that a particular item  $j$  may have a unique effect on the score  $Y$  by a degree of  $i_{0j}$ . Using the multilevel specification, this model is

$$\begin{aligned}y_{ij} &= \beta_0 + \beta_i x_{i1} + \beta_j x_{j1} + \varepsilon_{ij} \\ \beta_0 &= \gamma_0 + I_{0j} \\ \beta_i &= \gamma_{1.} \\ \beta_j &= \gamma_{.1}\end{aligned}\quad (5)$$

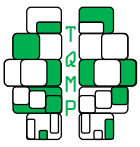
The SAS syntax to fit this *random item intercept model* is

```
PROC MIXED data=Data1 method=REML ic;
CLASS SubID ItemID;
MODEL y = X_i X_j /solution;
RANDOM intercept /subject=ItemID;
RUN;
```

The random intercept effects for items ( $I_{0j}$ ) is specified in the **RANDOM** statement with the option set to be **subject=ItemID**.

In addition, both subjects and items can have simultaneous systematic effects on the outcome variable through the specification of  $\beta_0 = \gamma_0 + S_{0i} + I_{0j}$ , which means that the outcome variable  $Y$  receives a unique effect from a particular subject answering to a specific item. Specifically, both **RANDOM** statements for the random intercept effect of subjects and items need to be included in the PROC MIXED procedure:

```
PROC MIXED data=Data1 method=REML ic;
CLASS SubID ItemID;
```



```
MODEL y = X_i X_j /solution;
RANDOM intercept /subject=SubID;
RANDOM intercept /subject=ItemID;
RUN;
```

### Random intercept and slope models

In addition to effects that can be attributed to either subject or item covariates (e.g., from gender or experimental manipulations), other systematic variations may exist in data. To account for otherwise unexplained individual subject or item variation, random slope effects can be included. Random slope effects are specified in the model by parameters  $\beta_i$  and  $\beta_j$  in Equation (2) for the effects of the subjects and items, respectively. The term "slope" can be somewhat confusing in the context of LMER. "Random slope effects" can be conceptualized as meaning that individual subjects (or items) behave differently when other factors are equal. That is, random effects on the slope parameters of  $\beta_i$  and  $\beta_j$  allow the slope of covariates  $x_{i1}$  and  $x_{j1}$  to vary in a systematic fashion. Like the random intercept effects, the random slope effects are assumed to be normally distributed with a mean of zero and a variance of  $U_{1i}^2$  for subjects and  $V_{1j}^2$  for items.

To formally define the random slope effects, let  $S_{1i}$  and  $I_{1j}$  denote the random variables representing the *random slope effects for subjects* and the *random slope effects for items*. The random slope effects for subjects and items can be included in the model individually or simultaneously. A model that allows the response variable,  $Y$ , to differ due to systematic variations in the data (random intercepts) and individual differences (random slopes) for subject covariates is

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_i x_{i1} + \beta_j x_{j1} + \varepsilon_{ij} \\ \beta_0 &= \gamma_0 + S_{0i} + I_{0j} \\ \beta_i &= \gamma_{1.} + S_{1i} \\ \beta_j &= \gamma_{.1} \end{aligned} \quad (6)$$

A model that allows the response variable,  $Y$ , to differ due to systematic variations in the data (random intercepts) and individual differences (random slopes) for item covariates is

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_i x_{i1} + \beta_j x_{j1} + \varepsilon_{ij} \\ \beta_0 &= \gamma_0 + S_{0i} + I_{0j} \\ \beta_i &= \gamma_{1.} \\ \beta_j &= \gamma_{.1} + I_{1j} \end{aligned} \quad (7)$$

Moreover, a model that has random intercepts effects and random slopes for both subject and item covariates to the subject and item covariate respectively is

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_i x_{i1} + \beta_j x_{j1} + \varepsilon_{ij} \\ \beta_0 &= \gamma_0 + S_{0i} + I_{0j} \\ \beta_i &= \gamma_{1.} + S_{1i} \\ \beta_j &= \gamma_{.1} + I_{1j} \end{aligned} \quad (8)$$

Since the model specified in Equations (6)–(8) has both random intercept and random slope effects, the covariance between the random intercept and random slope effects must also be included in the model and are assumed to have the following form:

$$\begin{pmatrix} S_{0i} \\ S_{1i} \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} U_{0i}^2 & U_{01,i} \\ U_{01,i} & U_{1i}^2 \end{bmatrix} \right) \quad (9a)$$

and

$$\begin{pmatrix} I_{0j} \\ I_{1j} \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_{0j}^2 & V_{01,j} \\ V_{01,j} & V_{1j}^2 \end{bmatrix} \right) \quad (9b)$$

where  $U_{01,i}$  and  $V_{01,j}$  are the covariances between the random intercept and random slope effects for the subjects and items, respectively. The variances and covariances between the subjects and items are assumed to be independent, and all the random effects are also assumed to be independent of residuals.

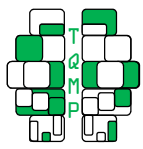
The *random slope effects* for subjects/items are specified by adding the subject/item covariate variables ( $X_i, X_j$ ) to the corresponding random statements. The three sets of PROC MIXED syntax to fit the models in Equations (6)–(8) are as follows.

For the model in Equation (6):

```
PROC MIXED data=Data1 method=REML ic;
CLASS SubID ItemID;
MODEL y = X_i X_j /solution;
RANDOM intercept X_i /subject=SubID type=UN;
RANDOM intercept /subject=ItemID;
RUN;
```

For the model in Equation (7):

```
PROC MIXED data=Data1 method=REML ic;
CLASS SubID ItemID;
MODEL y = X_i X_j /solution;
RANDOM intercept /subject=SubID;
RANDOM intercept X_j /subject=ItemID
```



```
type=UN;  
RUN;
```

For the random intercept and random slope effects for the subjects and items in Equation (8):

```
PROC MIXED data=Data1 method=REML ic;  
CLASS SubID ItemID;  
MODEL y = X_i X_j /solution;  
RANDOM intercept X_i /subject=SubID type=UN;  
RANDOM intercept X_j /subject=ItemID  
type=UN;  
RUN;
```

The most complex model is to include both random subject and item intercept effect, and random slopes effects for both subject and item covariate simultaneously is

$$\begin{aligned}y_{ij} &= \beta_0 + \beta_i x_{i1} + \beta_j x_{j1} + \varepsilon_{ij} \\ \beta_0 &= \gamma_0 + S_{0i} + I_{0j} \\ \beta_i &= \gamma_{1.} + S_{1i} + I_{1j} \\ \beta_j &= \gamma_{.1} + S_{1i} + I_{1j}\end{aligned}\quad (10)$$

The PROC MIXED syntax to fit this model in Equation (10) is

```
PROC MIXED data=Data1 method=REML ic;  
CLASS SubID ItemID;  
MODEL y = X_i X_j /solution;  
RANDOM intercept X_i X_j /subject=SubID  
type=UN;  
RANDOM intercept X_i X_j /subject=ItemID  
type=UN;  
RUN;
```

The system of specification equations presented above is similar to the two-level specification in a multilevel modeling framework. That is, in the current specification, the level 1 model specifies a regression model for the dependent variable explained by the covariates in this level. The parameters of the level 1 model are specified at level 2 by including covariates and/or random effects. As the subject and item covariates are in the same level, the current structure is usually referred to as a crossed random effects model (Raudenbush, 1993), which is different from the traditional nested structure usually found in multilevel modeling (where the subject and item covariates are in different levels).

In sum, the specifications of the LMER in the current

paper explicitly separate the sources of the subject and item effects (subject and item covariates) and also the types of effects (fixed or random) with dedicated terms so that models can be specified with different combinations of random intercept or random slope effects on subject or item covariates. These specifications can be easily specified in the PROC MIXED statements. As a result, this framework allows researchers to more easily represent and specify a wide range of effects in an LMER<sup>2</sup>.

### The SAS output

The SAS output for the previous specified models will include several parts. First of all, an “Estimation Iteration History” table is given to describe the iteration history. Second, the “Model Fitting Information” table shows general information of fit statistics (e.g., LL, AIC, BIC, -2LL, ...). Next, a table labeled “Covariance Parameter Estimates (REML)” is given which contains parameter estimates for all random components in the model. Finally, due to the specification of the `solution` option in the `MODEL` statement, a table “Solution for Fixed Effects” is presented, which contains the parameter estimates, estimated standard errors, and *t*-tests for all fixed effects in the model.

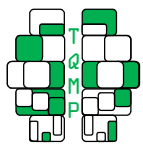
As has been discussed above, the different types of random effects in LMERS serve the context of psycholinguistic studies well. An LMER is an intuitive and flexible tool for model conceptualizations and specifications in data analysis. The second part of this paper will focus on the issues related to model selection under the framework of LMERS, with special attention paid to the random component.

### Model Selection for LMER Models

Model selection is an important step of any statistical analysis (for a general introduction to model selection see Hélie, 2006). In the context of LMERS, the task of model selection amounts to the selection of relevant explanatory variables (and interactions) in the fixed component and the structure of the random intercept and slope (with their covariance patterns) in the random component. The methods of model selection are usually discussed and studied separately for fixed effects and random effects; however, these two

---

<sup>2</sup> Sample SAS Code and toy dataset for the illustrated models are available upon request from the author.



components are related to each other as a change in the fixed effects component may affect the structure of the random component. This dependency between the fixed and random components also affects the estimations of parameters and therefore the statistical tests and predictions (Verbeke & Molenberghs, 2000). Consequently, the dependency between the fixed and random component should also be taken into consideration during the model selection process.

A survey of journal articles suggests that most applications that have employed LMERS did not provide theoretical or empirical justifications for the decision about how the random structure would be determined in the final model. One main reason for this is because, to date, the process for determining the "best fit" random structure has not been well studied and, in terms of random components, the general guidelines for model selection still need to be formulated.

The decision about the random structure in the LMER models is related to the role of the random effects. For most applications to psycholinguistic data, the "random component" is meant to account for or allow variability for subjects and/or items. By introducing the random effects, the dependency and variations in data can be better "modeled," "expressed," or "explained." For empirical data, we do not know the "true" structure of the random component; the best practice is to try to find the "best one" among the available candidate structures.

The importance of considering fixed and random components jointly is well acknowledged and supported. In this paper, the discussion is limited to the structure of the random component, which has not been studied extensively, especially with regard to its psycholinguistic data applications. The next section discusses two general approaches to deciding the structure of the random component of an LMER—hypothesis testing and using information criteria.

### **Hypothesis Testing**

Likelihood ratio tests (LRTs) are commonly used when selecting between two nested models (Hox, 2002; Snijders & Bosker, 1999). The major advantage of an LRT is that it is not affected by the parameterization of the model (Hox, 2002). More specifically, the accuracy of the LRT does not depend on the estimates of the variances and their standard errors. However, the hypotheses testing approach may not be a desired framework for model selection for some applications (Burnham & Anderson, 2002, p. 36). For example, one

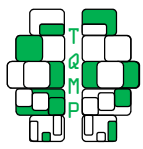
would not use this approach for an LMER to select the random structure of the model because in this context, model selection is used to test whether the variance (or covariance) terms are equal to zero. Self and Liang (1987) refer to this type of hypothesis as a nonstandard testing condition because the parameters of the null hypothesis fall at the boundary of the parameter space. Based on Self and Liang's (1987) results, Stram and Lee (1994, 1995) show that the asymptotic null distribution for the LRT statistics for testing a hypothesis of this type often consist of a mixture of chi-square distributions rather than the classical single chi-squared distribution. Under more general conditions, for example, comparing models with  $k$  and  $k+k'$  ( $k' > 1$ )  $\chi^2$  random effects, the null distribution is a mixture of random variables, and the weights for each component can only be calculated analytically in a number of special cases (Raubertas, Lee, & Nordheim, 1986; Shapiro, 1988).

In the context of mixed effects modeling, several model selection methods using the hypothesis testing approach have been proposed. For example, Raudenbush and Bryk (2002) use chi-squared tests for random intercepts and slopes; Stoel, Garre, Dolen, and van den Wittenboer (2006) discuss the use of this mixture of chi-squared approaches in a more general context using the chi-bar-square  $\bar{\chi}^2$  distribution; and Verbeke and Molenberghs (2003) propose using the score test to test the variance components. Although the LRT was also suggested for LMERS (Baayen et al., 2008; Pinheiro & Bates, 2000, p. 83), there are no details as to whether corrections or adjustments to the test statistics should be made in order to test null hypotheses with boundary problems.

It is important to note that when examining changes in model fit where the models differ in their random effects, the models need to be estimated using restricted maximum likelihood (REML). This is different when testing the significance of fixed effects where the parameters need to be estimated using maximum likelihood (ML). To test the significance of the random effects, the `covtest` option is included as a PROC MIXED option and is used to generate hypothesis testing output for variance and covariance components.

### **Information Criteria**

Comparing the values of information criteria (IC) among candidate models is a common method of model selection for non-nested models. The IC are based on the log-likelihood value. A likelihood function describes

**Table 2** ■ Summary of Information Criteria

| Criterion | Definition               | Reference               |
|-----------|--------------------------|-------------------------|
| AIC       | $-2LL + 2p$              | Akaike (1973)           |
| AICC      | $-2LL + 2pN/(N - p - 1)$ | Hurvich and Tsai (1989) |
| CAIC      | $-2LL + p(1 + \log N)$   | Bozdogan (1987)         |
| BIC       | $-2LL + p \log N$        | Schwarz (1978)          |
| HQIC      | $-2LL + 2p \log \log N$  | Hannan and Quinn (1979) |

the probability of observing the sampled data. The main idea behind IC is to compare models based on their maximized log-likelihood value and to penalize for the use of too many parameters. Specifically, the idea is to arrive at a model that describes the observed data to a satisfactory extent but without unnecessary complications. This approach is consistent with the general preference of parsimonious theories and models.

IC take the general form of

$$IC = -2 \log L + C \quad (11)$$

where the  $-2 \log L$  represents the degree of inaccuracy, the badness of the fit, or the bias when the ML estimates of the parameters of a model are used (Bozdogan, 1987, p. 356). The information criterion shown in Equation (11) suggests that, in addition to the badness of fit embedded in  $-2 \log L$ , the fit also penalizes for model complexity. The amount of penalty relating to model complexity is denoted as  $C$ . Therefore, the IC aim to find a good balance between model fitness (trying to maximize the likelihood function) and parsimony (penalizing additional complexity). It should be strongly emphasized that IC only provide rules of thumb to discriminate between several statistical models; they should never be used or interpreted as formal statistical tests of significance.

The Akaike Information Criterion (AIC) (Akaike, 1973) was one of the earlier IC propositions for model selection. AIC prefers a model that minimizes Equation (11) with  $C = 2p$ , where  $p$  is the number of parameters in the model. Corrected Akaike Information Criterion (AICC) is a finite sample corrected version of AIC proposed by Hurvich and Tsai (1989). The consistent AIC (CAIC), where  $C = p(1 + \log N)$ , was proposed by Bozdogan (1987). CAIC has the property of selecting models with fewer parameters than AIC.

Deriving from a different theoretical background, the Bayesian Information Criterion (BIC) (Schwarz,

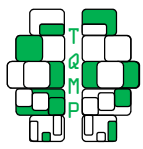
1978) is also a common criterion used in model selection. Using the formulation in Equation (11), the penalty term for the BIC is  $C = p \log N$ . In addition, the Hannan-Quinn information criterion (HQIC) proposed by Hannan and Quinn (1979) is available in the output of the PROC MIXED statement. These IC are provided in the output when the `ic` option is included in the PROC MIXED options statement.

The IC discussed are summarized in Table 2. As can be seen, AIC depends only on the number of parameters, while other measures depend on both the number of parameters and the sample size. The penalty term of BIC is more stringent than the penalty term of AIC for sample sizes larger than eight ( $p \log N$  exceeds  $2p$ ). Consequently, the BIC tends to favor smaller models compared to the AIC. Note that when applying IC in an LMER context, additional attention is required for the determination of "N". Specifically, when using ML as the estimation method for fitting the model, the total sample (N) should be used as "N" in computing IC, while  $N - p$  should be used when using the restricted maximum likelihood method (REML) in the estimation (Verbeke & Molenberghs, 2000, p.75). It is also important to know that different statistical software may use different "N" and "p" when calculating IC, so values provided in the standard output need to be used carefully. Further discussion of this can be found in Gurka (2006).

#### **Remarks about Model Selection in LMER**

In terms of model selection, hypothesis testing and the use of IC methods offer two different approaches for choosing the best model. The hypothesis testing approach usually employs sequential tests to arrive at the final model. On the other hand, the IC approach proceeds by comparing the fit for several (if not all) candidate models and selecting the best fit among those choices. This approach is more in line with the "simultaneous" concept of model selection. However, other approaches have been suggested for LMER. For example, Wolfinger (1992) and Diggle, Liang, and Zeger





(1994) have suggested a two-step approach where an over-fitted fixed effects model (the most complex mean structure) is fitted using ML, and then the covariance structures (random effects) are selected using REML estimation by IC. Verbeke and Molenberghs (2000) propose an iterative approach in which the procedure iterates between mean structure (fixed effect) and covariance structure (random effects) or even residual covariance structure. They also encourage the use of graphs and plots to guide the selection process. For example, typical plots are the ordinary least squares (OLS) residual profile and the smoothed average trend of the squared OLS residuals.

The decisions made in the model selection process can be either substantive-driven or data-driven. With a substantive-driven decision, researchers may base their decision on field knowledge, existing theory, detailed problem formation, and common sense in forming a specific structure of the random component. One may wish to test whether such a structure is appropriate or supported by the collected data. On the other hand, a data-driven decision when selecting random components is not guided by specific theories or hypotheses. A data-driven decision in model selection aims to search for what the random structure is, based on the collected data. A common practice for this approach is to propose several candidate models and have each fitted to the data. The "best" one is then identified according to a chosen criterion.

No mechanical procedures for evaluating model fit should override human judgment (Brown, 2000). Similarly, no criteria or tests should be used blindly; considerations such as experimental designs, research goals, or substantive theories should be taken into account in evaluating model fit. For example, researchers need to decide whether the focus is on fixed or random effects. In most cases, fixed effects are the primary concern of studies; therefore, the appropriate random structure is desired to provide more accurate parameter estimations for fixed effects. However, researchers might want to include all relevant fixed effects in a model when random effects are the focus.

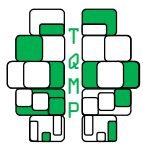
## Discussion and Conclusion

In this paper, an alternative multilevel specification of LMERS in the application of psycholinguistic data was described. This paper also illustrated how these models can be specified and estimated using PROC MIXED in SAS. Although previous papers have illustrated the

LMERS using R (e.g., Baayen et al, 2008) and SPSS (e.g., Carson & Beeson, 2013), no paper has used the two-level specification approach to systematically illustrate how different models can be specified using PROC MIXED in SAS. The only exception is Locker et al. (2007), who took the multilevel approach but only focused on the item analysis aspect, also using SAS PROC MIXED. The illustrated multilevel specifications in this paper separate the item and subject factors (covariates) for the fixed and random components. As a result, the effects deriving from different sources can be clearly represented and distinguished. The distinction between the item and subject covariates is especially helpful in selecting the appropriate fixed and random effects structures in the context of psycholinguistic studies. For example, different effects are specified in the "random" statement of the PROC MIXED in SAS.

The issues relating to model selection in the context of LMERS were also discussed. Specifically, whether the focus should be on fixed or random effects and whether the decision should be either substantive- or data-driven. Given one data set, a researcher may come up with more than one model, each of which can be an apparently satisfactory model according to a specific model selection procedure or criterion. This reflects the basic indeterminacy that is inherent in model fitting based on empirical data. It is quite possible to have several different models but no compelling argument in favor of any particular one. In such cases, it is better to accept this indeterminacy and leave it to future research to resolve than to make an unwarranted decision. There is no universal standard guideline or rule governing model selection, but this review and discussion of the concept of model selection could offer new insights for researchers when analyzing empirical data.

The theoretical development of statistical tests and methods has consistently advanced the understanding of model selection; however, the practitioners and applied researchers who use these methods do not typically keep track of new developments in quantitative methodologies. Many developments require substantial statistical knowledge and programming skills in order to apply the new methods to empirical studies. Therefore, the newly developed methods are usually not very user-friendly for applied researchers in terms of empirical applications. Thus, this paper is helpful for researchers who wish to apply the methods using standard statistical packages, such as SAS. This paper offers a bridge to connect the gap

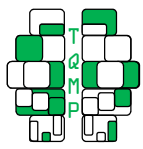


between method development and applying LMERS to empirical research.

In summary, LMERS have many desirable properties in analyzing psycholinguistic data. Comprehensive statistical knowledge is essential when applying LMERS to ensure accurate inferences and conclusions. This paper not only illustrated how PROC MIXED in SAS can be used to specify various LMER models but also provided detailed information about selecting the proper structure of the random component. The discussion and illustration in this paper aim to help researchers apply LMERS in the field of psycholinguistics both in model specification and on the issue of model selection.

## References

- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*, 255–265. doi: 10.2307/2334537
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370. doi: 10.1007/BF02294361
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer-Verlag.
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*, 108–132. doi: 10.1006/jmps.1999.1279
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 135–142. doi: 10.1016/S0022-5371(73)80014-3
- Carson, R. J., & Beeson, C. M. L. (2013). Crossing language barriers: Using crossed random effects modelling in psycholinguistics research. *Tutorials in Quantitative Methods for Psychology*, *9*, 25–41.
- Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. New York: Oxford University Press.
- Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician*, *60*, 19–26. doi: 10.1198/000313006X90396
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society B*, *41*, 190–195. Retrieved from <http://www.jstor.org/stable/2985032>
- Hélie, S. (2006). An Introduction to Model Selection: Tools and Algorithms. *Tutorials in Quantitative Methods for Psychology*, *2*, 1–10.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
- Hurvich, C. M., & Tsai, C. -L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307. doi: 10.1093/biomet/76.2.297
- Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*, 42–73. doi: 10.1016/j.jml.2011.03.002
- Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to item analysis of psycholinguistic data. *Behavior Research Methods*, *39*, 723–730. doi: 10.3758/bf03192962
- Moreton, E. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language*, *67*, 165–183. doi: 10.1016/j.jml.2011.12.003
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York: Springer.
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425. doi: 10.1016/j.jml.2008.02.002
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Raubertas, R. F., Lee, C. I. C., & Nordheim, E. V. (1986). Hypothesis tests for normal means constrained by linear inequalities. *Communications in Statistics – Theory and Methods*, *15*, 2809–2833. doi: 10.1080/03610928608829280
- Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross sectional and longitudinal research. *Journal of Educational Statistics*, *18*, 321–349. doi: 10.3102/10769986018004321
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, Second Edition. Newbury Park, CA: Sage.



- Rouder, J. N. (2005). Applicability of deadline models: Comment on Glickman, Gray, and Morales. *Psychometrika*, *70*, 427–430. doi: 10.1007/s11336-004-1215-4
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. doi:10.1214/aos/1176344136
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343. doi:10.1007/BF02294360
- Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*, 605–610. doi:10.1080/01621459.1987.10478472
- Snijders, T. A. B., & Bosker, R. J. (1999). *An introduction to basic and advanced multilevel modeling*. London: Sage.
- Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, *60*, 308–327. doi:10.1016/j.jml.2008.11.002
- Stoel, R. D., Garre, F. G., Dolen, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, *11*, 439–455. doi: 10.1037/1082-989X.11.4.439
- Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*, 1171–1177. doi: 10.2307/2533455
- Stram, D. O., & Lee, J. W. (1995). Correction to "Variance components testing in the longitudinal mixed effects model." *Biometrics*, *51*, 1196. doi: 10.2307/2533038
- Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, *56*, 49–62. doi: 10.2307/1403361
- Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer Series in Statistics. New York: Springer-Verlag.
- Verbeke, G., & Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, *59*, 254–262. doi:10.1111/1541-0420.00032
- Wolfinger, R. (1992). *A tutorial on mixed models*. Cary, NC: SAS Institute Inc.

## Citation

Yu, H.-T. (2015). Applying Linear Mixed Effects Models with Crossed Random Effects to Psycholinguistic Data: Multilevel Specification and Model Selection. *The Quantitative Methods for Psychology*, *11* (2), 78-88.

Copyright © 2015 Yu. This is an open-access article distributed under the terms of the *Creative Commons Attribution License (CC BY)*. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 23/09/14 ~ Accepted: 14/03/15