Title
Modern Applications of Cross-classified Multilevel Models (CCMMs) in Social and Behavioral Research: Illustrations with R Package PLmixed

Author
Huang, Sijia

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Modern Applications of Cross-classified Multilevel Models (CCMMs) in Social and Behavioral Research: Illustrations with R Package PLmixed

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Sijia Huang

2021

ABSTRACT OF THE THESIS

Modern Applications of Cross-classified Multilevel Models (CCMMs) in Social and Behavioral Research: Illustrations with R Package PLmixed

by

Sijia Huang
Master of Science in Statistics
University of California, Los Angeles, 2021
Professor Yingnian Wu, Chair

Respondents in social and behavioral studies often belong to two or more non-nested higher-level groups of aggregation simultaneously, yielding the so-called cross-classified data structure. For example, in education, students belong to the schools they attend and the neighborhoods they live in, and there exists no exact nesting between the schools and neighborhoods. The cross-classified multilevel model (CCMM; Goldstein, 1994; Rasbash & Goldstein, 1994) was introduced as an extension of the standard multilevel model to accommodate the prevalent cross-classified data. The CCMM has been mainly applied in education to study the impacts of various contexts on certain outcomes, such as the influence of schools and neighborhoods on smoking behaviors among adolescents (Dunn, Richmond, Milliren, & Subramanian, 2015). However, applications of the CCMM in other fields are relatively scant and little-known. One potential reason for this lack of applications could be the limited availability of software programs that allow the easy fit of the CCMM.

To advocate more applications of the CCMM in a broader spectrum, in this article, we first show the connections between the CCMM and several widely used psychometric models, including the random effect item response theory (IRT) model (Van den Noortgate, De Boeck, & Meulders, 2003), the model for rater effects (e.g., Murphy & Beretvas,

2015), the multitrait-multimethod (MTMM) model (Campbell & Fiske, 1959), and the generalizability theory (G-theory) model (Shavelson & Webb, 1991). Then we review a few modern applications of the CCMM, such as its applications to meta-analyses and social network analysis (SNA).

To address the issue of software programs, we introduce a flexible and efficient R package **PLmixed** (Jeon & Rockwood, 2017), and show how the above-mentioned related models and applications of the CCMM can be estimated with **PLmixed** and other existing R packages. Finally, we conclude that the CCMM would be applied more broadly with the support of computer software such as **PLmixed**.

The thesis of Sijia Huang is approved.

Chad Hazlett

Minjeong Jeon

Yingnian Wu, Committee Chair

University of California, Los Angeles

2021

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

This chapter starts with a review of the prevalent multilevel data structures in social and behavioral sciences in order to facilitate the understanding of the cross-classified multilevel model (CCMM; Goldstein, 1994; Rasbash & Goldstein, 1994). Then, we elaborate the issue regarding applications the CCMM and discuss our research goals.

1.1 Background

The respondents in social and behavioral studies often belong to one or more higher-level groups of aggregation. Thus, data in these studies often possess multiple levels (Bryk & Raudenbush, 1992). The ubiquitous multilevel data can be either hierarchical, or have non-hierarchical structures, such as the cross-classified or the multiple membership structures. To take into account the multilevel structures while analyzing data, several multilevel models, including the standard multilevel model (Bryk & Raudenbush, 1992), the CCMM and the multiple membership model (Hill & Goldstein, 1998) have been developed.

An essential feature of the hierarchical data is that each lower-level unit belongs to one and only one higher-level unit. An example of the three-level hierarchical data in education is that students are nested within classrooms, and classrooms are nested within schools. As shown in the unit diagram (Figure 1.1a), each student (level 1) belongs to one and only one classroom (level 2), and each classroom belongs to one and only one school (level 3). Another example of the hierarchical data is that in family studies, family members such as parents and children are the level-1 units, and they are considered to be nested within the families at a higher level (level 2). In addition, data collected in longitudinal studies, where respondents are measured at multiple time points, are also hierarchical. The measurements (level 1) can be viewed as being nested within respondents (level 2).

In cross-classified data, there exists no such exact nesting as in the hierarchical data. Instead, lower-level units are nested within combinations of higher-level units defined by two or more classifications. For example, an shown in Figure 1.1b, students (level 1) belong to schools (level 2) and neighborhoods (level 2) simultaneously. The schools are not nested within neighborhoods (e.g., school sc1 does not belong to any of the neighborhoods), and vice versa. The students are cross-classified by schools and neighborhoods. Another example of the cross-classified data arises in clinical studies where patients are nested within their residential areas and, at the same time, their clinicians. Since patients who live in the same area do not typically go to the same clinician and patients of the same clinician do not live in the same area, residential areas and clinicians are not nested within one another. Thus, the patients (level 1) are considered to be nested within the cross-classifications of their residential areas and clinicians (i.e., two crossed factors at level 2).

When lower-level units belongs to more than one higher-level units of the same classification, the data possess the so-called multiple membership structure. For example, in Figure 1.1c, the student s2 moves between schools and belongs to both schools sc1 and sc2. Outcomes of students such as the academic performance are influenced by two schools.

The assumption of independent observations that underlies models for single-level data (e.g., regression models) is violated in the context of multilevel data analyses. For example, in a longitudinal study in which the respondents' loneliness levels are measured at multiple time points, the measurements of the same respondent would not be independent, and would be more similar than those of different respondents. Thus, the multilevel data structures have to be taken into consideration while analyzing multilevel data.

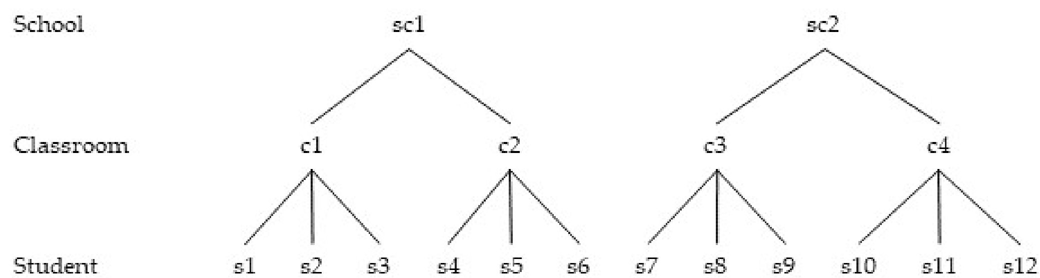
To account for the lack of independence in hierarchical data, the standard multilevel model, also known as the hierarchical linear model (Bryk & Raudenbush, 1992), was developed and has been broadly applied in education and psychology. The cross-classified multilevel model (CCMM; Goldstein, 1994; Rasbash & Goldstein, 1994) and the multiple membership model (Hill & Goldstein, 1998) have also been proposed as extensions of the standard multilevel model to accommodate the cross-classified and multiple membership data structures. Browne, Goldstein, and Rasbash (2001) synthesized these two extensions and introduced the multiple membership multiple classification (MMMC) model, in which lower-level units are cross-classified by higher-level units of two or more classifications and can belong to more than one higher-level units of the same classification. In Chapter 2, the general model formulations of the CCMM and the MMMC are introduced.

1.2 Research Goal

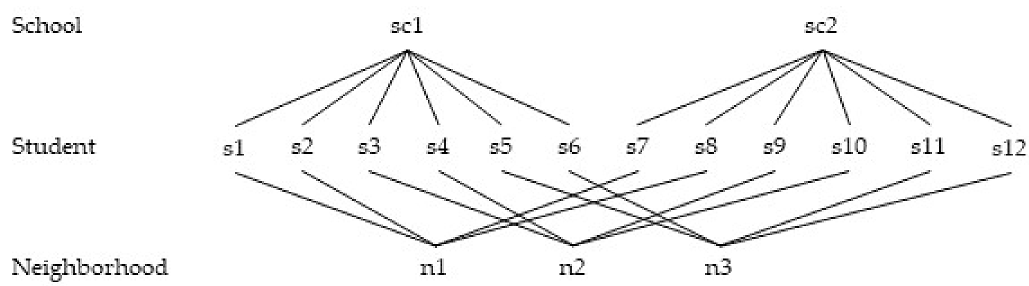
It is worth noting here that in addition to the above-mentioned examples of cross-classified data, data in various studies in social and behavioral sciences also possess cross-classified structures. For example, item response data can be viewed as cross-classified data, where the item responses are cross-classified by persons and items. Although data with cross-classified structures are prevalent, applications of the CCMM are relatively scant, and are mostly in the field of education (e.g., Goldstein, Burgess, & McConnell, 2007; Leckie, 2009; Rasbash, Leckie, Pillinger, & Jenkins, 2010). The CCMM has been mainly applied to study the impacts of various contexts on certain outcomes such as students' academic achievement.

One potential reason for this lack of applications is the limited availability of software programs that allow users to easily fit the CCMM. To the best of our knowledge, `MLwiN` (Charlton, Rasbash, Browne, Healy, & Cameron, 2020) is the most widespread software package for multilevel analyses. However, fitting a CCMM with `MLwiN` is not straightforward. Although the Markov Chain Monte Carlo (MCMC) estimation procedure is available in `MLwiN`, it could be somewhat challenging for users who are not familiar with the Bayesian modeling approach. For example, improper priors would introduce biases to model parameter estimates, resulting in misleading interpretations. Another option for fitting the CCMM is the `R` (R Core Team, 2017) package **lme4** (Bates, Mächler, Bolker, & Walker, 2015). However, **lme4** cannot estimate models with unknown factor loadings or item discrimination parameters.

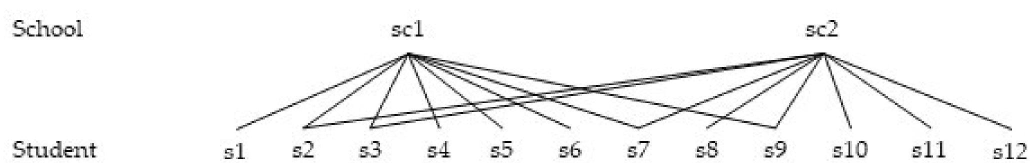
The goal of the present study is two-fold. First, we present how several measurement models and methods that are broadly used in psychology are connected with the CCMM, including the random effect item response theory (IRT) model, the model for rater effects, the multitrait-multimethod (MTMM) model, and the generalizability theory (G-theory) model. We also introduce a few modern applications of the CCMM, including its applications in meta-analyses, social network analysis (SNA), and contextual effects. Second, to address the issue of software programs, we introduce a flexible and efficient `R` package for estimating models with multiple hierarchical levels and/or crossed random effects, **PLmixed** (Jeon & Rockwood, 2017, 2018), and demonstrate how the CCMM can be estimated with **PLmixed** and other existing `R` packages in various contexts.



(a) A Unit Diagram for Three-level Hierarchical Data



(b) A Unit Diagram for Cross-classified Data



(c) A Unit Diagram for Multiple Membership Data

Figure 1.1: Unit Diagrams for Multilevel Data

CHAPTER 2

Model Formulation and Software Programs

In this chapter, we first present the general formulation of the CCMM in an educational context. Then we review existing software programs for the estimation of the CCMM, and introduce the functionality of the `R` package **PLmixed** (Jeon & Rockwood, 2017, 2018).

2.1 Model Formulation

To set up the notations, let's consider a study in which researchers would like to study the effects of schools and neighborhoods on students' math performance. The math performance can either be students' scores of a standardized test or binary evaluations of pass of fail. The researchers also want to explore how characteristics of schools (e.g., if the school is a public school), neighborhoods (e.g., residential income deprivation score), and students (e.g., student age, ethnicity) are related to the outcome measure. There is no nesting between schools and neighborhoods, because students attend the same school can live in different neighborhoods and students live in the same neighborhood can attend different schools. Thus, the level-1 units, students, are cross-classified by two level-2 crossed factors, school and neighborhoods.

Let $y_{i(jk)}$ denote the math performance of student i's $(i = 1, \ldots, I)$ who goes to school j $(j = 1, \ldots, J)$ and lives in neighborhood k $(k = 1, \ldots, K)$. I, J and K are the total numbers of students, schools and neighborhoods, respectively. We can specify the model below,

$$g\left[y_{i(jk)}\right] = \mathbf{X}_{i(jk)}\boldsymbol{\beta} + u_j + u_k + \varepsilon_{i(jk)} \tag{2.1}$$

where $g(\cdot)$ is a link function that transforms the linear predictor on the right-hand side to the conditional expectation of the response variable $y_{i(jk)}$. Using a link function allows accommodating outcome variables of various distribution families. For example, if $y_{i(jk)}$ is continuous and assumed to be normally distributed (e.g., scores of standardized tests), an identity link function can be used; and if $y_{i(jk)}$ is binary (e.g., pass/fail), a probit link function can be applied so that the linear predictor (which ranges from $-\infty$ to ∞) is mapped to a 0 to 1 scale.

On the right-hand side of Equation (2.1), $\mathbf{X}_{i(jk)}$ is a covariate matrix, each row of which corresponds a student and each column of which corresponds to a covariate. $\boldsymbol{\beta}$ is a vector of fixed effects or regression coefficients. The first column of $\mathbf{X}_{i(jk)}$ is usually $\mathbf{1}$ so that the first element of $\boldsymbol{\beta}$ represents the intercept. If we have $y_{i(jk)}$ as the math score, the intercept can be interpreted as the grand mean and represents the math score of an *"average"* student in an *"average"* school and an *"average"* neighborhood. If no covariates were included in the model, Equation (2.1) would reduce to the so-called *"intercept-only model"*,

$$g\left[y_{i(jk)}\right] = \boldsymbol{\beta}_0 + u_j + u_k + \varepsilon_{i(jk)} \tag{2.2}$$

In both Equations 2.1 and 2.2, u_j and u_k are respectively the random effects associated with the two level-2 crossed factors, schools and neighborhoods, and are assumed to be normally distributed, $u_j \sim \mathcal{N}(0, \sigma^2_{\text{sch}})$ and $u_k \sim \mathcal{N}(0, \sigma^2_{\text{nei}})$. The last term $\varepsilon_{i(jk)}$ represents the level-1 residual, with $\varepsilon_{i(jk)} \sim \mathcal{N}(0, \sigma^2_{\text{e}})$. Note here the parentheses are used to emphasize that the schools and neighborhoods are at the same level. For more details regarding the formulation and extensions of the cross-classified multilevel model, we refer interested readers to comprehensive review articles (e.g., Fielding & Goldstein, 2006; Cafri, Hedeker, & Aarons, 2015).

2.2 Software program

In this section, we review existing computer softwares for the estimation of the CCMM, and introduce the functionality of **PLmixed** (Jeon & Rockwood, 2017, 2018), which can be downloaded from CRAN (`https://cran.r-project.org/web/packages/PLmixed/`).

2.2.1 Existing packages

MLwiN. `MLwiN` (Charlton et al., 2020) is one of the most widely used statistical software packages for fitting multilevel models and has many advantages. For example, it can model various response types, including continuous, binary variables, and count. In addition, `MLwiN` can accommodate models with inconstant level-1 variance (i.e., heteroscedasticity). It implements the iterative generalized least squares (IGLS) and MCMC estimation procedures. However, `MLwiN` is not straightforward as long as the CCMM is concerned. As estimating the CCMM is usually computationally intensive, it is recommended that the users split the data into separate groups to make additional levels to reduce the storage needed. Although the MCMC estimation procedure is available in `MLwiN`, using the MCMC procedure requires some fundamental knowledge about the Bayesian modeling approach. For example, users have to be carefully about selecting priors for parameters, especially the priors for variances of random effects (which are of great importance in the CCMM). Inappropriate priors would result in biased parameter estimates and misleading interpretations.

lme4. The `R` package **lme4** (Bates et al., 2015) includes functions to fit various kinds of models with both fixed and random effects, including the linear, generalized linear and nonlinear mixed effect models. **lme4** can easily incorporate models with nested and crossed random effects, of which the CCMM is an example. It has also been applied to estimate the one parameter logistic (1PL) item response model (Doran, Bates, Bliese, Dowling, et al., 2007). However, **lme4** cannot estimate models with unknown factor loadings or item discrimination parameters.

2.2.2 `R` package PLmixed

The `R` package **PLmixed** (Jeon & Rockwood, 2017, 2018) extends the capabilities of **lme4** to estimate extended generalized linear mixed models with factor structures. **PLmixed** maintains all merits of **lme4** and requires minimum learning of new syntax for users who are already familiar with **lme4**, as the syntax structure of **PLmixed** follows directly from **lme4**. In terms of model estimation, **PLmixed** implements a profile maximum likelihood estimation approach (Jeon & Rabe-Hesketh, 2012) through using **lme4** and the `optim` function (Byrd, Lu, Nocedal, & Zhu, 1995).

To demonstrate the major advantage of **PLmixed**, let's consider a standard generalized linear mixed model (GLMM), which takes the form,

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\delta. \tag{2.3}$$

In Equation (2.3), $g(\cdot)$ is the link function, $\boldsymbol{\mu}$ is the conditional expectation of the response variable, $\mathbf{X}$ and $\mathbf{Z}$ are respectively the covariate matrices associated the vectors of fixed effects $\boldsymbol{\beta}$ and random effects δ. Equation (2.3) can be extended by introducing a factor structure $\boldsymbol{\Lambda}$ so that

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\Lambda}\boldsymbol{\eta}, \tag{2.4}$$

where $\boldsymbol{\eta}$ is the random effect vector, with $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, $\mathbf{W}$ is the corresponding covariate matrix, and $\boldsymbol{\Lambda}$ is the matrix of factor loadings. If all elements in $\boldsymbol{\Lambda}$ are fixed to 1, Equation (2.4) becomes Equation (2.3). Popular measurement models in psychological and educational studies, including factor analysis and item response theory models, can be formulated as an extended GLMM described with Equation (2.4). **PLmixed** allows the factor loading matrix $\boldsymbol{\Lambda}$ to be freely estimated (which is not feasible with **lme4**), leading to more flexibility in model fitting.

The main function of the **PLmixed** package is `PLmixed`. Main arguments of the `PLmixed` function that are identical to the `glmer` function of **lme4**: `formula` (a two-sided linear formula that describes the model), `data` (a data frame containing variables in `formula`), and `family` (a GLMM family to specify the distribution of μ). To estimate factor loadings in the extended GLMM, three new syntax commands are introduced: `load.var` (a vector of variables correspond to `lambda` and `factor`), `lambda` (the factor loading matrix), and `factor` (a list of factor names in `lambda`). We will further illustrate the usage of the `PLmixed` function in the following chapter through multiple applications of the CCMM in various contexts.

CHAPTER 3

Models in Psychometrics as CCMMs

Many measurement models and methods that are broadly used in psychology can be viewed as CCMMs. In this chapter, we present the connections between the CCMM and several popular psychometric model, including the random effect item response theory (IRT) model (Van den Noortgate et al., 2003), the model for rater effect, the multitrait-multimethod (MTMM) model (Campbell & Fiske, 1959), and the generalizability theory (G-theory) model (Shavelson & Webb, 1991). We also show how these models can be fitted with **PLmixed**. For each model, we first provide a brief summary of related studies, then describe a sample data and the model formulation, and finally analyze the data using **PLmixed**.

3.1 Random Effects Item Response Theory (IRT) Model

Item response theory (IRT) models aim to describe the relationship among categorical item responses, psychometric properties of items, and the latent variable(s) theorized to influence the item responses. A classic unidimensional IRT model for items scored at two categories (e.g., 1/0 as correct/incorrect response to an test item, present/not present a symptom) is the Rasch model (Rasch, 1960),

$$P(x_{ij} = 1 \mid \theta) = \frac{\exp(\theta_j - \delta_i)}{1 + \exp(\theta_j - \delta_i)}. \tag{3.1}$$

where θ_j is the theorized latent variable, and is often assumed to follow a standard normal distribution, $\theta_j \sim \mathcal{N}(0, 1)$. δ_i is an item parameter and indicates item i's difficulty. Equation 3.1 describes how the conditional probability that person j's $(j = 1, \ldots, J)$

response to item i $(i = 1, \ldots, I)$ equals 1, $P(x_{ij} = 1 \mid \theta)$, changes as a function of θ_j and δ_i. This classic IRT model assumes random person effect and fixed item effect.

Van den Noortgate et al. (2003) proposed a new type of IRT models by combining the CCMM with the traditional IRT model. This new type of IRT models assume that both persons and items are random samples from the person and item populations so that both person and item effects are random. The item responses (level 1) are nested within the cross-classifications of persons (level 2) and items (level 2). For an item with two score categories, the proposed random item effect IRT model is,

$$\text{logit}(\pi_{ij}) = \beta_0 + u_{1j} + u_{2i} \tag{3.2}$$

where π_{ij} denotes the probability that person j endorses item i, β_0 is the logit of π_{ij} of an *average* person on an *average* item (i.e., $u_{1j} = u_{2i} = 0$), or the difference between the overall item easiness and the overall person trait levels. u_{1j} represents person j's latent trait level, and is assumed to follow a normal distribution with mean 0 and standard deviation σ_{u1}, $u_{1j} \sim \mathcal{N}(0, \sigma^2_{u1})$. u_{2i} indicates item i's easiness level, and is also assumed to be normally distributed, $u_{2i} \sim \mathcal{N}(0, \sigma^2_{u2})$. This random item effect model allows researchers to incorporate person- and item- characteristics as covariates, and including additional levels such as a school level.

We analyze a simulated dataset in the **PLmixed** package to show how an IRT problem is approached with the cross-classified model. The simulated dataset `IRTsim` contains responses of 500 students from 26 schools to 5 dichotomously-scored items. We install the **PLmixed** package to print out the first six rows.

```
> install.packages("PLmixed")
> library(PLmixed)
>
> data(IRTsim)
> head(IRTsim)
    sid school item y
1.1   1      1    1 1
1.2   1      1    2 1
1.3   1      1    3 1
```

```
1.4   1        1     4 0
1.5   1        1     5 1
2.1   2        1     1 1
```

The outcome variable (`y`; 1: correct, 0: incorrect) is the response of a student (`sid`) in a school (`school`) to an item (`item`).

Model In this example, item responses (level 1) are cross-classified by items and students (level 2), and one of the level-2 factors, students, are nested within schools (level 3). Thus, we specify the following model:

$$\text{logit}\left[\pi_{ij(k)}\right] = \beta_0 + u_i + u_{j(k)} + u_k \tag{3.3}$$

where $\pi_{ij(k)}$ is the probability that the student j in school k correctly answer item i, and via the logit$(\cdot)$ function, the linear predictor on right-hand side of the equation ($-\infty$ to ∞) is mapped to the probability scale (0 to 1). β_0 is the overall intercept, and represents the baseline logit when $u_i = u_{j(k)} = u_k = 0$. u_i is the item random effect and indicates how easy an item is. $u_{j(k)}$, and u_k are student and school random effects, respectively. These random effects are assumed to follow normal distributions, $u_i \sim \mathcal{N}(0, \sigma^2_{\text{item}})$, $u_{j(k)} \sim \mathcal{N}(0, \sigma^2_{\text{stu}})$, $u_k \sim \mathcal{N}(0, \sigma^2_{\text{sch}})$. Note here the parentheses in the $u_{j(k)}$ term is used the emphasizes that students are nested within schools.

Fitting the model The model is fitted with the below **PLmixed** syntax,

```
> IRT.example <- PLmixed(formula = y ~ 1 + (1|item) + (1|sid:school)
+                                      + (1|school),
+                           data = IRTsim, family = binomial)
```

The argument `formula` follows Equation (3.3): `y` is the binary outcome variable, `1` represents the overall intercept of the linear predictor at the right-hand side of Equation (3.3) and corresponds to β_0, and `(1|item)`, `(1|sid:school)` and `(1|school)` are the three random-effect terms. The colon in the term `(1|sid:school)` indicates that the students are nested within schools. Here, we use the argument `family = binomial` so that the binomial family with a logit link function is applied. This corresponds to the logit$(\cdot)$ on the left side of Equation (3.3). If not specified, the default identity link function would be used. We save all results in an object named `IRT.example`.

After fitting the model, we can print all the results through the `summary()` function.

```
> summary(IRT.example)
Profile-based Mixed Effect Model Fit With PLmixed Using lme4
Formula:  y ~ 1 + (1 | item) + (1 | sid:school) + (1 | school)
Data:  IRTsim
Family: binomial  ( logit )

     AIC      BIC   logLik deviance df.resid
 2981.94  3005.24 -1486.97  2407.03     2496

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2215 -0.9390  0.5634  0.8253  1.9891

Random effects:
 Groups     Name        Variance Std.Dev.
 sid:school (Intercept) 0.8322   0.9122
 school     (Intercept) 0.7293   0.8540
 item       (Intercept) 0.1592   0.3990
Number of obs: 2500, groups:  sid:school, 500; school, 26; item, 5

Fixed effects:
              Beta     SE z value Pr(>|z|)
(Intercept) 0.6966 0.2545   2.737 0.006192

lme4 Optimizer:  bobyqa
Optim Optimizer:  NA
Optim Iterations:  1
Estimation Time:  0.02 minutes
```

The first section of results echos the formula, the dataset that contains variable in the formula and the link function. The second section of results presents model fit indices including AIC, BIC, and the residual summary. Then in the `Random effects` section, the estimates of variances of random effects and the associated standard errors are presented. These parameter estimates indicate that within a school, the students' math proficiency follows a normal distribution with variance 0.83, the school average math proficiency follows a normal distribution with variance 0.73, and the item easiness parameter follows a normal distribution with variance 0.16. Numbers of observations, students, schools and items are shown in this section as well. Following the `Random effects` section, in the `Fixed effects` section, the estimate of the fixed effect (the overall intercept in our case) is presented. An estimate of about 0.70 indicates that the probability of the an *average* student (i.e., $u_{j(k)} = 0$) from an *average* school (i.e., $u_k = 0$) correctly answers an *average* item (i.e., $u_i = 0$) is about 0.69 ($1/[1 + \exp(-0.70)]$). The last section of results lists `lme4` optimizer, `optim` optimizer, the number of `optim` iterations, and the estimation time.

3.2 Modeling Rater Effect

Many research questions in psychology and education uses ratings provided by raters. These ratings are inevitably subject to the rater bias. The CCMM has been applied to both cross-sectional (e.g., Murphy & Beretvas, 2015; Lei, Li, & Leroux, 2018; Jayasinghe, Marsh, & Bond, 2003) and longitudinal data (e.g., Guo & Bollen, 2013) to model rater effects. For example, in Jayasinghe et al. (2003), the impact of rater characteristics (e.g., gender) and grant proposal characteristics on ratings of these proposals were studied using the cross-classfied multilevel model.

We simulated a dataset that mimic the empirical dataset introduced in Jayasinghe et al. (2003). The simulated dataset consists of 2,401 evaluations of 1,580 assessors on 673 grant proposals. Each assessor and proposal belong to one of the 28 fields. We first print out the first six rows of the simulated dataset.

```
> head(RATERsim)
  field      rating assessor assessor_gender proposal researcher_gender
1     1  0.75122876       19               0      446                 0
2     1 -0.30709805       20               0      579                 0
3     1 -0.03705511       20               0      184                 0
4     1 -0.40030331       20               0       71                 0
5     1 -0.46344476       28               1      117                 0
6     1  0.43064563       34               0      293                 1
```

The outcome variable is the rating (`rating`) of a proposal (`proposal`) provided by an assessor (`assessor`) in a field (`field`). The mean and standard deviation of the ratings are 0.14 and 1.05, respectively. The number of proposals assessed by each assessor ranges from 1 to 3. The number of ratings each proposal receives ranges from 2 to 7. The `field` variable indicates the field the ratings belong to. The numbers of assessors, proposals and ratings in the fields range from 15 to 83, from 12 to 33, and from 37 to 122, respectively. Other covariates included in the simulated dataset are genders of the assessor (`assessor_gender`) and first author of the proposal (`researcher_gender`). For these two covariates, 0 indicates the assessor/author is male, and 1 indicates the assessor/author is a female.

Model Since each assessor evaluates more than one proposals and each proposal receives more than one ratings, the ratings (level 1) are cross-classified by assessors and proposal (level 2). These two level-2 factors are both nested within fields (level 3). In addition to the effects of assessor and author genders on ratings, another interesting question to ask could be how genders interact with ratings (e.g., if female assessors rate proposals written by female authors consistently higher). To answer these research questions, the model for the simulated rating data can be specified as,

$$y_{ij(k)} = \beta_0 + \beta_1 x_{ik} + \beta_2 x_{jk} + \beta_3 x_{ij(k)} + u_{i(k)} + u_{j(k)} + u_k + \varepsilon_{ij(k)} \tag{3.4}$$

where $y_{ij(k)}$ is the rating of proposal j provided by assessor i in field k . $x_{i(k)}$, $x_{j(k)}$ are respectively the gender of assessor i and the first author of proposal j. To address the gender-rating interaction question, a variable $x_{ij(k)}$ is generated, and takes the value of 1 if the assessor and the first author are both females and 0 otherwise. Note that $x_{i(k)}$ and $x_{j(k)}$ are covariates of level-2 units, while $x_{ij(k)}$ is associated with lelve-1 units. β_0 to β_4 are the intercept and regression coefficients associated with the three covariates. $u_{i(k)}$, $u_{j(k)}$ and $u_{ij(k)}$ are the assessor, proposal and field random effects, respectively, and are assumed to be normally distributed, $u_{i(k)} \sim \mathcal{N}(0, \sigma^2_{\text{assessor}})$, $u_{j(k)} \sim \mathcal{N}(0, \sigma^2_{\text{proposal}})$, $u_k \sim \mathcal{N}(0, \sigma^2_{\text{field}})$. $\varepsilon_{ij(k)}$ is the residual term and is also assumed to follow a normal distribution, $\varepsilon_{ij(k)} \sim \mathcal{N}(0, \sigma^2_e)$.

Fitting the model The below PLmixed syntax is used to fit the model specified in Equation (3.4),

```
> rater.example <- PLmixed(formula = rating ~ 1 + assessor_gender
+                          + researcher_gender
+                          + assessor_gender*researcher_gender
+                          + (1|assessor:field) + (1|proposal:field)
+                          + (1|field),
+                          data = RATERsim, family = gaussian)
```

In the syntax, `rating` represents the ratings in the simulated data. The `assessor_gender`, `researcher_gender`, and `assessor_gender*researcher_gender` are the three covariate terms. The `(1|assessor:field)`, `(1|proposal:field)` and `(1|field)` are the three random effect terms. In this example, the default identify link is used through the argument `family = gaussian`.

After fitting the model, we use the `summary()` function to obtain the estimation results. To save space, we present here only parameter estimates and the associated standard errors in the `Random effects` and `Fixed effects` sections.

```
> summary(rater.example)
Random effects:
 Groups         Name        Variance  Std.Dev.
 assessor:field (Intercept) 8.876e-01 0.942151
 proposal:field (Intercept) 1.672e-01 0.408946
 field          (Intercept) 2.541e-02 0.159413
 Residual                   9.575e-06 0.003094
Number of obs: 2401, groups:  assessor:field, 1580; proposal:field, 673;
field, 28

Fixed effects:
                                      Beta       SE t value
(Intercept)                        0.12239 0.043287   2.827
assessor_gender                   -0.15918 0.080937  -1.967
researcher_gender                  0.14543 0.042842   3.395
assessor_gender:researcher_gender -0.03682 0.002183 -16.866
```

In the `Random effects` section, estimates of variances and the corresponding standard deviations are shown. As consistent with results in Jayasinghe et al. (2003), assessors explain 82% of the total variation in ratings (0.8876/[0.8876+0.1672+0.02541]), proposals explain about 15% of the variation, and fields explain 3% of the variation. In the `Fixed effects` section, estimates of regression coefficients and the associated standard errors are presented. All estimates are significant at a 95% level, indicating that female assessors on average provide lower ratings (-0.16) than male assessors, female authors on average receive higher ratings (0.15) than male authors, and proposals written by female authors receive lower ratings (-0.04) from female assessors.

3.3 Multitriat-Multimethod Model

The multitrait-multimethod (MTMM; Campbell & Fiske, 1959) model allows researchers to establish convergent and discriminant validity between traits and investigate effects of methods. The MTMM approach has been applied in psychology to study personality (e.g., Biesanz & West, 2004; DeYoung, 2006) and life satisfaction (e.g., Lance & Sloan, 1993). Typical MTMM data consist of measures of multiple traits (e.g., the Big Five) obtained through multiple methods (e.g., self report or peer report). These (often continuous) measures can be viewed as being cross-classified by traits and methods, since each trait is evaluated with multiple methods and each method is applied to measure more than one traits. Jeon and Rijmen (2014) applied the Monte Carlo local likelihood (MCLL; Jeon, Kaufman, & Rabe-Hesketh, 2019) algorithm to categorical MTMM data and provide an empirical illustration.

To show the connection between the MTMM appraoch and the CCMM, we simulated a dataset that reproduces the MTMM covariance matrix reported in DeYoung (2006). The simulated data include ratings (in a 5-point Likert scale) on the Big Five obtained from subjects ($n = 500$) and three of their peers. For each subject, the five self-reported ratings are treated as responses to item 1-5, and the 15 peer-reported (five for each peer) ratings are treated as responses to separate items (labeled as item 6-10). A diagram is shown in Figure 3.1, in which the circles are the variables (E = Extraversion, A = Agreeableness, C

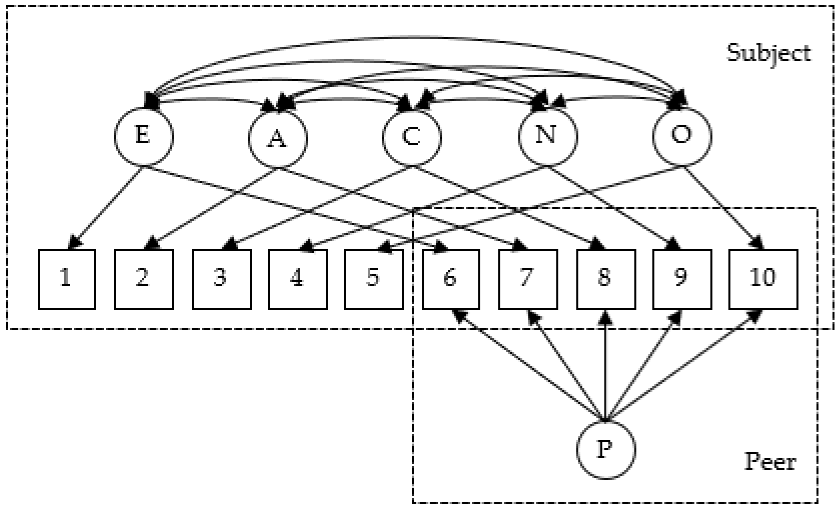


Figure 3.1: Diagram of the Multitriat-Multimethod Model

= Conscientiousness, N = Neuroticism, O = Openness/Intellect, P = peer report effect) and the squares are observed ratings (item 1-5 are self-reported ratings on five latent traits, item 6-10 are peer-reported).

The first six row of the simulated data are shown below.

```
> head(MTMM.data)
     subject item method trait peer    score
[1,]       1    1      1     1    0 4.464043
[2,]       1    2      1     2    0 3.619863
[3,]       1    3      1     3    0 4.301094
[4,]       1    4      1     4    0 4.301457
[5,]       1    5      1     5    0 3.121121
[6,]       1    6      2     1    1 5.131464
```

The variable `subject`, `item`, `method`, `trait` and `peer` are indicators of subject, item, method (1: self report, 2: peer report), trait (1 = Extraversion, 2 = Agreeableness, 3 = Conscientiousness, 4 = Neuroticism, 5 = Openness/Intellect) and peer (0: self report). The last column, `score`, is the outcome variable.

Model The model for the data can be specified as,

$$y_{itpj} = \beta_i + \lambda_{it}^T \theta_{tj}^T + \lambda_{ip}^M \theta_p^M + \varepsilon_{itpj} \tag{3.5}$$

where y_{itpj} represents peer p's rating to item i on trait t of subject j. β_i denotes the intercept of item i, λ^T_{it} and λ^M_{ip} are item i's loading on θ^T_{tj}, the latent variable that corresponds to trait t, and θ^M_p, the rater effect latent variable, respectively. θ^T_{tj} is subject j's level in trait t, and it is assumed that all five trait latent variables follow a multivariate normal distribution $\boldsymbol{\theta}^T_j \sim \mathcal{N}(0, \boldsymbol{\Sigma}_T)$. θ^M_p represents peer p's effect on the items, $\theta^M_p \sim \mathcal{N}(0, \sigma^2_M)$. ε_{itpj} denotes the residual term, $\varepsilon_{itpj} \sim \mathcal{N}(0, \sigma^2_e)$.

Fitting the model To fit the model, we have to specify the factor loading `lambda` matrix according to the diagram (Figure 3.1) first. Rows of the `lambda` matrix correspond to items, and columns correspond to latent variables. For identification purpose, we fix some loadings to be 1. Note here that as only item 6-10 are peer-reported items, the first five elements of the rater effect latent variable (i.e., the last column of the `lambda` matrix) are zeros. `NA` represents unknown parameters that are to be estimated.

```
> lambda  <- rbind(c(1, 0, 0, 0, 0, 0),
+                  c(0, 1, 0, 0, 0, 0),
+                  c(0, 0, 1, 0, 0, 0),
+                  c(0, 0, 0, 1, 0, 0),
+                  c(0, 0, 0, 0, 1, 0),
+                  c(NA,0, 0, 0, 0, 1),
+                  c(0,NA, 0, 0, 0, NA),
+                  c(0, 0,NA, 0, 0, NA),
+                  c(0, 0, 0,NA, 0, NA),
+                  c(0, 0, 0, 0,NA, NA))
```

The full `PLmixed` syntax can be specified as the following:

```
> MTMM.example <- PLmixed(score ~ 0 + as.factor(item)
+                        + (0+E+A+C+N+O|subject)
+                        + (0+rater|peer),
+                        data = as.data.frame(MTMM.data),
+                        lambda = list(lambda), load.var = "item",
+                        factor = list(c("E","A","C","N","O",
+                                        "rater")))
```

The `0` is included in the `formula` argument to avoid estimating extra random intercepts. We use `as.factor(item)` so that item-specific intercepts are estimated. The five trait latent variables are named `E`, `A`, `C`, `N` and `O`. As assumed, these latent variables are correlated and vary across subjects. If independence is assumed among the trait latent variables, we can use `(E|subject) + (A|subject) + (C|subject) + (N|subject) + (O|subject)`. The last term in the `formula` argument corresponds to the random rater effects. The `lambda` argument reads in the factor loading matrix we specified earlier. The `load.var` argument indicates the variable that defines the factor loading matrix. And the `factor` shows the names of the latent variables.

After fitting the model, we can use the `summary` function to extract the results.

```
>summary(MTMM.example)
Profile-based Mixed Effect Model Fit With PLmixed Using lme4
Formula: score~0+as.factor(item)+(0+E+A+C+N+O|subject)+(0+rater|peer)
Data:  as.data.frame(MTMM.data)
Family: gaussian  ( identity )

     AIC      BIC   logLik deviance df.resid
19181.24 19440.82 -9554.62 19109.24     9964

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3027 -0.5599 -0.0007  0.5554  3.0555

Lambda:  item
       E    SE     A    SE     C    SE    N   SE      O    SE rater    SE
1  1.000     .     .     .     .     .    .    .      .     .     .     .
2      .     . 1.000     .     .     .    .    .      .     .     .     .
3      .     .     .     . 1.000     .    .    .      .     .     .     .
4      .     .     .     .     .     . 1.00    .      .     .     .     .
5      .     .     .     .     .     .    .    .  1.000     .     .     .
6  0.905 0.035     .     .     .     .    .    .      .     . 1.000     .
7      .     . 0.992 0.075     .     .    .    .      .     . 0.913 0.064
8      .     .     .     . 0.829 0.044    .    .      .     . 0.828 0.042
9      .     .     .     .     .     . 1.84 0.14      .     . 1.536 0.120
10     .     .     .     .     .     .    .    . -2.406 0.177 1.164 0.097

Random effects:
 Groups   Name  Variance Std.Dev. Corr
 peer     rater 0.06228  0.2496
 subject  E     0.46781  0.6840
          A     0.15870  0.3984    0.11
```

```
         C     0.18075  0.4251    0.11  0.12
         N     0.40576  0.6370   -0.10 -0.44 -0.28
         O     0.31660  0.5627    0.27  0.01  0.08 -0.16
 Residual      0.21202  0.4605
Number of obs: 10000, groups:  peer, 1501; subject, 500

Fixed effects:
                   Beta      SE t value
as.factor(item)1  3.336 0.03687   90.48
as.factor(item)2  4.064 0.02723  149.23
as.factor(item)3  4.060 0.02803  144.86
as.factor(item)4  2.572 0.03515   73.18
as.factor(item)5  3.670 0.03252  112.87
as.factor(item)6  3.670 0.03081  119.12
as.factor(item)7  4.104 0.02438  168.31
as.factor(item)8  4.185 0.02324  180.06
as.factor(item)9  2.610 0.03065   85.17
as.factor(item)10 3.734 0.02513  148.58

lme4 Optimizer:  bobyqa
Optim Optimizer:  L-BFGS-B
Optim Iterations:  1120
Estimation Time:  36.61 minutes
```

The first section of the results echos the formula, and the second section shows model fit indices. The `Lambda` section includes estimates of factor laodings and the corresponding standard errors. Following that in the `Random effects` section, estimates of variances of all six latent variables and the corresponding standard errors are presented. The estimated correlation between the trait latent variables are shown as well. In the `Fixed effects` section, item-specific intercepts are presented. In the last section, information about `lme4` optimizer is shown.

3.4 Generalizability Theory Model

Generalizability theory (G-theory) is a statistical framework for evaluating the generalizability (reliability) of behavioral measurements (Shavelson & Webb, 1991). In G-theory, observed scores of a measurement are decomposed into additive effects of multiple facets (i.e., sources of variations). If levels of a facet in a measurement design are viewed as random samples from the universe of all possible levels and researchers would like to generalize beyond these observed levels, this facet is considered as a *random* facet. In contrast, if all possible levels a facet are included in the design or there is no need for generalization, this facet is a *fixed* facet. G-theory includes two types of studies, the generalizability study (G study) and the decision study (D study). The aim of a G study is to computed the variances associated with the facets. In a D study, different generalizability coefficients are to be constructed, using the variance estimates from the G study. A generalizability coefficient is analogous to a reliability coefficient but depends on which facets are considered random.

Vangeneugden, Laenen, Geys, Renard, and Molenberghs (2005) adopted the idea of G-theory and utilized a flexible linear mixed effect model framework to derive the generalizability/reliability coefficients. They applied the proposed approach to individual patient data of five double-blind randomized clinical trials to compute different generalizability coefficients of the Positive and Negative Syndrome (PANSS).

To demonstrate the connection between generalizability theory studies and the CCMM, we analyze the `Brennan.3.2` dataset that can be found in the **gtheory** package. The dataset contains scores of 10 persons' performance on 3 tasks, each of which is rated by 4 raters (i.e., the classical person*[rater:task] design). In this data set, the ratings can be viewed as being cross-classified by persons and raters, while the raters are nested within tasks.

We first print the first six rows of the data.

```
> install.packages("gtheory")
> library(gtheory)
>
> head(Brennan.3.2)
  Task Person Rater Score
1    1      1     1     5
2    1      2     1     9
3    1      3     1     3
4    1      4     1     7
```

```
5     1         5        1        9
6     1         6        1        3
```

The outcome variable is the score (`Score`), which ranges from 1 to 9. The 10 persons (`Person`) complete all three tasks (`Task`). Raters (`Rater`) 1 to 4 assess task 1, raters 5 to 8 assess task 2, and raters 9 to 12 assess task 3.

Model In this simple person*(rater:task) design, there are five sources of variations: persons, tasks, raters (nested within tasks), person-task interactions and residuals. We can specify the following model:

$$y_{ij(k)} = \beta_0 + u_i + u_{j(k)} + u_k + u_{ik} + \varepsilon_{ij(k)} \tag{3.6}$$

where $y_{ij(k)}$ is person i's score on task k by rater j. Equation (3.6) can be viewed as a cross-classified multilevel model, where u_i represents the person random effect, $u_{j(k)}$ is the rater random effect, u_k is the task random effect, u_{ik} is the person-task random effect, and $\varepsilon_{ij(k)}$ is the residual term associated with each observation. Each of random effect and residual terms is assumed to follow a normal distribution, $u_i \sim \mathcal{N}(0, \sigma^2_{\text{person}})$, $u_{j(k)} \sim \mathcal{N}(0, \sigma^2_{\text{rater}})$, $u_k \sim \mathcal{N}(0, \sigma^2_{\text{task}})$, $u_{ik} \sim \mathcal{N}(0, \sigma^2_{\text{person*task}})$, and $\varepsilon_{ij(k)} \sim \mathcal{N}(0, \sigma^2_{\text{e}})$.

Fitting the model The `PLmixed` syntax and `Random effects` and `Fixed effects` sections of the results are shown below,

```
> g.example <- PLmixed(formula = Score ~ 1 + (1|Person) + (1|Rater:Task)
+                       + (1|Task) + (1|Person:Task),
+                       data = Brennan.3.2, family = gaussian)

> summary(g.example)
Random effects:
 Groups      Name        Variance Std.Dev.
 Person:Task (Intercept) 0.5730   0.7570
 Rater:Task  (Intercept) 0.6475   0.8047
 Person      (Intercept) 0.4290   0.6550
 Task        (Intercept) 0.1561   0.3951
 Residual                2.3802   1.5428
Number of obs: 120, groups:  Person:Task, 30; Rater:Task, 12; Person, 10;
Task, 3
```

```
Fixed effects:
            Beta     SE t value
(Intercept) 4.75 0.4334   10.96
```

With these parameter estimates, we can derive multiple generalizability/reliability coefficients, which are often desired in G-theory studies. For example, if we want to generalize persons' score over raters and tasks, we can construct the following generalizability coefficient,

$$
\begin{aligned}
R_{\rho^2\text{Rel}} &= \frac{\sigma^2_{\text{person}}}{\sigma^2_{\text{person}} + \sigma^2_{\text{rater}} + \sigma^2_{\text{task}} + \sigma^2_{\text{person*task}} + \sigma^2_{\text{e}}} \\
&= \frac{0.5730}{0.5730 + 0.6475 + 0.4290 + 0.1561 + 2.3802} = 0.14
\end{aligned}
$$

CHAPTER 4

Modern Applications of CCMM

In this chapter, we introduce a few modern and recent applications of the CCMM and demonstrate these applications with `R` packages, including the CCMM's applications to meta analysis, social network analysis (SNA), and to the study of impacts of contexts on educational achievement of immigrant students. In addition, we introduce the formulation of the multiple membership multiple classification (MMMC) model (Browne et al., 2001), which can be viewed as a general form of the CCMM. We also show an application of the MMMC to SNA.

4.1 Meta Analysis

Meta-analysis is a systematic approach that combines results from multiple primary studies that address the same research question and is able to derive more statistically powerful findings. As pointed out in Hox, Moerbeek, and Van de Schoot (2017), the meta-analysis can be viewed as special case of multilevel analysis, where the quantities of interest (e.g., the effect size; ES) are nested within primary studies. However, often each primary study in a meta-analysis calculate more than one ESs based on the same sample, and at the same time, multiple primary studies report ESs of the same outcome, creating additional dependency between ESs. Fernández-Castilla et al. (2019) illustrated the use of cross-classified multilevel model to deal with the dependency between ESs caused by multiple crossed random factors (e.g., studies and outcomes). Through simulations, they found that the cross-classified modeling approach yields unbiased parameter estimates in the context of meta-analysis, outperforming the multilevel and multivariate approaches.

To illustrate the application of the CCMM to meta-analysis, we use a subset of data of the *Meta-Analytic Study of Loneliness* (MASLO; Maes et al., 2017) project. The data used can be also found in Appendix A of Fernández-Castilla et al. (2019). This dataset includes 68 standardized mean gender differences in loneliness, reported from 54 studies on seven possible subscales, including the UCLA Loneliness Scale (coded as 1) and the Social dimension of the SELSA (coded as 2). The first six rows of the dataset are shown below.

```
> head(meta)
  Study Outcome Subscale      g Variance Precision
1     1       1        1 -0.251    0.024    41.455
2     2       1        1 -0.069    0.001  1361.067
3     3       1        5  0.138    0.001   957.620
4     4       1        1 -0.754    0.085    11.809
5     5       1        1 -0.228    0.020    49.598
6     6       1        6 -0.212    0.004   246.180
```

The outcome variable that of interest is standardized gender difference in loneliness (g), the `Variance` shows the associated sampling variance, the variable `Precision` varible is the reciprocal of the corresponding `Variance`. `Study` is the study identification number, `Outcome` indicates the outcome reported across studies, and `Subscale` is the tool used in the study to measure loneliness.

Model The meta-analysis data possesses a three-level data, which can be specified as the following

$$y_{i(jk)} = \beta_0 + u_j + u_k + u_{i(jk)} + \varepsilon_{ijk} \tag{4.1}$$

where $y_{i(jk)}$ is the ith outcome that is measured with subscale j in the kth study – that is to say, the outcomes (level 2) are nested within the cross-classifications of subscales and studies (level 3). ε_{ijk} represents the known level-1 sampling variance. The intercept β_0 is the pooled gender difference.

Fitting the model The model specified in Equation 4.1 can be fitted using the `blmer` function in the **blme** package (Dorie & Dorie, 2015). The **blme** package extends the **lme4**

package and allows the estimation of linear and generalized linear mixed-effects models in a Bayesian setting.

```
> library(blme)
> meta.example <- blmer(g ~ 1 + (1|Study) + (1|Subscale)
+                          + (1|Outcome:Study:Subscale),
+                      data=meta, weights = Precision,
+                      resid.prior = point(1), cov.prior=NULL)
```

in which the `formula` argument corresponds to the model. The two `blmer` specific arguments `resid.prior = point(1)` and `weights = Precision` fix the level-1 variance to 1 and assigning weights to each observation so that we have ε_{ijk}^2 fixed to the value of `Variance` associated with each observation.

After fitting the model, we use the `summary` function and present estimates of the random and fixed effects.

```
> summary(meta.example)
Random effects:
 Groups                 Name        Variance Std.Dev.
 Outcome:Study:Subscale (Intercept) 0.009446 0.09719
 Study                  (Intercept) 0.011897 0.10907
 Subscale               (Intercept) 0.036287 0.19049
 Residual                           1.000000 1.00000
Number of obs: 68, groups:  Outcome:Study:Subscale, 68; Study, 57;
Subscale, 7

Fixed effects:
            Estimate Std. Error t value
(Intercept) -0.03809    0.08370  -0.455
```

Estimates in the `Random effects` section indicate there is more variations among subscales (0.04) than among studies (0.01). As shown in the `Fixed effects` section, the estimate of the combined gender difference is -0.04, and the corresponding t-value is -0.46, meaning that the gender difference is not significantly different from zero.

4.2 Social Network Analysis

Social network analysis (SNA) is a new but rapidly growing field. SNA is a strategy for studying social structures and how structural regularities would influence individual behaviours (Otte & Rousseau, 2002). Tranmer, Steel, and Browne (2014) included the network configurations (e.g., ego-nets, cliques of size 2 and 3) as a type of classification in a CCMM so that the social network dependence is accounted for. They found that ignoring the network would lead to biased estimates to both fixed and random parts of the model. Koster, Leckie, Miller, and Hames (2015) present a multilevel formulation of the Social Relations Model (SRM) to address the endogeneity problem in dyadic network data analysis. This multilevel SRM can be formulated as a CCMM whose outcome follows a Poisson distribution under reasonable assumptions.

De Nooy (2011) applied the cross-classified model to longitudinal social network analysis and developed a multilevel discrete-time event history model for time-stamped longitudinal network data. He illustrated the approach using a data set that contains all reviews and interviews among 40 most frequently-appearing literary authors and critics in the Netherlands, 1970–1980, in which the book reviews are cross-classified by authors (*heads*) and critics (*tails*). He used the discrete-time hazard (the probability of an event happening in a time period conditional on no such event happened in earlier time periods) as the dependent variable and estimated the model with multilevel logistic regression.

To illustrate the application of the CCMM in the context of social network analysis, we analyze a dyadic network dataset that can be found in supplementary materials of Koster et al. (2015). The dataset was collected in a village of indigenous Ye'kwana horticulturalists with eight households in Venezuela. These eight house household lead to 28 dyads. We first order the dataset by `dyad`, and then print out the first six rows (three dyads).

```
> network <- network[order(network$dyad),]
> head(network)
```

	giver	receiver	dyad	relationship	sharing	association	distance	kinship
1	1	2	12	1	1	0.234	162	0.0156
8	2	1	12	8	0	0.233	162	0.0160
2	1	3	13	2	1	0.144	286	0.0000
15	3	1	13	15	5	0.144	286	0.0000
3	1	4	14	3	0	0.120	327	0.0000
22	4	1	14	22	2	0.120	327	0.0000

The outcome variable is the total number of meals (`sharing`) provided from one household (`giver`) to another (`receiver`). Each giver-receiver pair defines a relationship (`relationship`). Note the `relationship` variable is directed – it is not the same for household A-B and B-A pairs. Other covariates included in the dataset are the distance between households (`distance`), the genetic relatedness between households (`kinship`), and a association index that provides measure of interactions between each pair of households (`association`).

Model The model for the dyadic network data can be specified as:

$$g(\pi_{ij}) = \beta_0 + \beta_1 x_{1|ij|} + \beta_2 x_{2|ij|} + \beta_3 x_{3|ij|} + g_i + r_j + u_{|ij|} + e_{ij} \tag{4.2}$$

where $g(\cdot)$ is a link function so that the number of meals household i provides to household j, y_{ij}, follows a Poisson distribution of mean π_{ij}, $y_{ij} \sim \text{Pois}(\pi_{ij})$. β_0 is the intercept in the linear predictor, and β_1, β_2, β_3 are regression coefficients associated with the relationship-level covariate, $x_{1|ij|}$, $x_{2|ij|}$, and $x_{3|ij|}$. g_i and r_j represent the giver and receiver random effects, respectively, and are assumed to follow normal distributions, $g_i \sim \mathcal{N}(0, \sigma^2_{\text{giver}})$, $r_j \sim \mathcal{N}(0, \sigma^2_{\text{receiver}})$. $u_{|ij|}$ is the dyad random effect, and is assumed to follow a normal distribution with mean 0 and standard deviation σ_{relation}. e_{ij} is the residual term, and $e_{ij} \sim \mathcal{N}(0, \sigma^2_e)$.

Fitting the model For simplicity, we fit the a CCMM with no covariates using the below `PLmixed` syntax,

```
> network.exp  <- PLmixed(sharing ~ 1 + (1|giver) + (1|receiver)
+                                 + (1|dyad) + (1|relationship),
+                           data = network, family = poisson)
```

Then we use the `summary()` function with the name of the object for the fitted model within the parentheses to obtain the output.

```
> summary(network.exp)
Random effects:
 Groups       Name        Variance Std.Dev.
 relationship (Intercept) 0.2401   0.4900
 dyad         (Intercept) 1.1990   1.0950
 receiver     (Intercept) 0.1060   0.3256
 giver        (Intercept) 1.8853   1.3731
Number of obs: 56, groups:  relationship, 56; dyad, 28; receiver, 8; giver, 8

Fixed effects:
              Beta     SE z value Pr(>|z|)
(Intercept) 0.3114 0.5668  0.5495   0.5827
```

With estimates of giver, receiver and relationship random effects, we can compute the variance partition coefficient (VPC) to quantify the contribution of each variation source to the total variance. For example, the VPC of giver is

$$VPC_{giver} = \frac{1.8853}{0.2401 + 1.1990 + 0.1060 + 1.8853} = 0.549$$

4.3 Immigration Studies

The CCMM has also been applied to study how multiple contexts help explain differences in educational achievement of immigrant students. We demonstrate how `PLmixed` can be used to disentangle the effects of origin country, destination country, and community on immigrant students' math achievement using a simulated data set that mimic the 2003 PISA survey (Levels, Dronkers, & Kraaykamp, 2008). The simulated data set includes observations of 7,403 immigrant students who were born in 35 different countries and took the test in 13 different countries. We first print out the first six rows of the simulated data set.

```
> head(round(PISAsim))
  student destination origin community score
```

```
1         1           1      6          1    429
2         2           1      6          1    479
3         3           1      6          1    569
4         4           1      6          1    428
5         5           1      6          1    596
6         6           1      6          1    436
```

The outcome variable is the math achievement (`score`) of a student (`student`) who was born in a country (`origin`) and took the PISA test in another country (`destination`). The mean and standard deviation of scores are 503 and 98, respectively. Each combination of the origin and the destination countries defines a immigrant community (`community`). Since not all origin countries are presented in all destination countries, the simulated dataset contains 67 (instead of $13 \times 35 = 455$) different immigrant communities. The number of students in the communities ranges from 90 to 144.

Model In this example, we have the students (level 1) nested within communities (level 2), and communities are cross-classified by countries of origin and destinations (level 3). To disentangle these contextual effects, we specify the following model with no covariate,

$$y_{i(jk)} = \beta_0 + u_j + u_k + u_{jk} + \varepsilon_{ijk} \tag{4.3}$$

where $y_{i(jk)}$ represents the math score of student i's, who was born in country j, took the test in country k and lives in the immigrant community defined by these two countries. β_0 is the overall intercept and can be interpreted as the grand mean of math scores. u_j and u_k are origin country and destination country random effects, and each is assumed to follow a normal distribution so that $u_j \sim \mathcal{N}(0, \sigma^2_{\text{ori}})$ and $u_k \sim \mathcal{N}(0, \sigma^2_{\text{des}})$. u_{jk} is an interaction term of the countries of origin and destination, and represents the community random effect, $u_{jk} \sim \mathcal{N}(0, \sigma^2_{\text{com}})$. ε_{ijk} is the residual term associated with student i in the community defined by the origin and destination countries of this student, and is assumed to follow a normal distribution, $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2_e)$.

Fitting the model The above model can be estimated with the syntax,

```
> PISA.example <- PLmixed(formula = score ~ 1 + (1|origin) + (1|destination)
+                          + (1|origin:destination), data = PISAsim)
```

Elements of the `formula` argument correspond to the terms in Equation (4.3). The `1` represents the overall intercept, and three terms `(1|origin)`, `(1|destination)` and `(1|origin:destination)` correspond to the three random-effect terms.

After fitting the model, we obtain the results through the `summary()` function.

```
> summary(PISA.example)
Profile-based Mixed Effect Model Fit With PLmixed Using lme4
Formula:  score ~ 1 + (1|origin) + (1|destination) + (1|origin:destination)
Data:  PISAsim
Family: gaussian  ( identity )

      AIC       BIC    logLik  deviance  df.resid
 87238.60  87273.15 -43614.30  87228.60      7398

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.2489 -0.6745  0.0057  0.6636  3.7325

Random effects:
 Groups             Name        Variance Std.Dev.
 origin:destination (Intercept)  511.2   22.61
 origin             (Intercept)  803.2   28.34
 destination        (Intercept)  681.4   26.10
 Residual                       7459.7   86.37
Number of obs: 7403, groups:  origin:destination, 67; origin, 35;
destination, 13

Fixed effects:
            Beta    SE t value
(Intercept)  500 9.478   52.75

lme4 Optimizer:  bobyqa
Optim Optimizer:  NA
Optim Iterations:  1
Estimation Time:  0.02 minutes
```

The first part of the results echos the model formulation. Then, the `Random effects` section shows estimates of variances of random effects and associated standard errors. Based on these estimates, the proportions of the total variation in math scores explained by the community students stay in, country of origin and destination country are 5.4%

(511.2/[511.2+803.2+681.4+7459.7]), 8.5% (803.2/[511.2+803.2+681.4+7459.7]) and 7.2% (681.4/[511.2+803.2+681.4+7459.7]), respectively. Following the `Random effects` section, the `Fixed effects` section presents the estimate of intercept and its standard error. The last section of the results summarizes the lme4 optimizer.

4.4 Multiple-Membership Multiple-Classification (MMMC) Model

Recall that the multiple membership model (Hill & Goldstein, 1998) was proposed to accommodate the multiple membership data structure, in which each lower unit can belong to more than one high-level units of the same classification (e.g., a student can move between schools). Browne et al. (2001) combined the CCMM with the multiple membership model and introduced the multiple membership multiple classification (MMMC) model.

4.4.1 Model Formulation

Before presenting the formulation of the MMMC model, we first introduce the multiple membership model. Let's consider an example of the multiple membership data in education. Students may change schools throughout their schooling, and their academic achievement are impacted by all schools they attend. For students who attend more than one school, it is natural to assume that the impacts of schools on them vary, since they spend different amount of time in these schools. In the multiple membership model, a weighting scheme is introduced,

$$\begin{gathered} y_i = \beta_0 + \sum_{j \in School(i)} w_{i,j} u_j + e_i \\ u_j \sim \mathcal{N}(0, \sigma_u^2) \quad \text{and} \quad e_i \sim \mathcal{N}(0, \sigma_e^2) \end{gathered} \tag{4.4}$$

In Equation 4.4, y_i represents student i's $(i = 1, \ldots, N)$ academic achievement, where N is the total number of students. β_0 is the intercept. u_j is the random effect of school j, $u_j \sim \mathcal{N}(0, \sigma_u^2)$, $w_{i,j}$ is the associated weight, and $\sum_{j \in School(i)} w_{i,j} = 1$. This term indicates that the impact of schools on student i's academic achievement is a weighted sum of effects of all schools this student attend. Usually, the weight $w_{i,j}$ is specified by the researchers. e_i is the residual term and is assumed to follow a normal distribution, $e_i \sim \mathcal{N}(0, \sigma_e^2)$.

Consider a more complicated case, in which students' academic performance are impacted by the neighborhoods they stay in (assuming no move between neighborhoods) and all schools they attend. Using the notation in Browne et al. (2001), the MMMC model distinguishes the two types of classifications (i.e., neighborhoods and schools) via superscripts and takes the following form,

$$y_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i^{(2)} \mathbf{u}_{C_{2(i)}}^{(2)} + \sum_{j \in C_3(i)} w_{i,j}^{(3)} \mathbf{Z}_i^{(3)} \mathbf{u}_j^{(3)} + e_i$$

$$\mathbf{u}_{C_{2(i)}}^{(2)} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{u(2)}), \quad \mathbf{u}_j^{(3)} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{u(3)}), \quad \text{and} \quad e_i \sim \mathcal{N}(0, \sigma_e^2). \tag{4.5}$$

where $\boldsymbol{\beta}$ is a vector of fixed effects. $\mathbf{u}_{C_{2(i)}}^{(2)}$ and $\mathbf{u}_j^{(3)}$ are respectively the random effects of the neighborhood student i stay in and school j, and are both assumed to be normally distributed. $w_{i,j}$ is the weight assigned to the school j that student i attend. $\mathbf{X}_i$, $\mathbf{Z}_i^{(2)}$ and $\mathbf{Z}_i^{(3)}$ are vectors of predictors. e_i is the residual term and is assumed to follow a normal distribution.

4.4.2 An Application

We use simulated data to illustrate how the MMMC model is fitted. The simulated data mimic the study reported in Tranmer et al. (2014) and consist of two parts: the academic performance of a sample of 968 (`N`) students in 10 schools, and a network adjacency matrix $\mathbf{D}$. The first six rows of the academic performance data are shown below, whose first column (`stu.id`) indicates students ids, the second column (`sch.id`) shows the schools the students are in, and the third column (`y`) are standardized test scores, $y \sim \mathcal{N}(0, 1)$.

```
> head(mmmc)
  stu.id sch.id           y
```

```
1        1        1 -1.03085195
2        2        2 -1.34750042
3        3        6 -0.96452530
4        4        2  0.48814972
5        5        8  0.03055814
6        6        8 -2.14955272
```

Each student in the sample can nominate up to ten friends, who can be either in the sample or out of the sample. The student who nominates friends is defined as *ego* in an "ego-net" and the nominated friends are defined as *alters*. This friendship network can be summarized using a 968×968 network adjacency matrix $\mathbf{D}$. The element in the i-row and j-th column of $\mathbf{D}$ takes the value of one if student i nominates student j as a friend and zero otherwise. For example, if student i nominates three students in the sample as friends, three elements in the i-th row of the $\mathbf{D}$ matrix are ones and all the other elements in the i-th row are zeros. The ego-net size of student i's is three. Among the 968 students in the sample, 651 (`N.ego`) students have non-zero ego-nets. The mean size of the non-zero ego-nets is 1.80 with a standard deviation of 1.22.

Model The students are cross-classified by the schools they are in and ego-nets of other students in the sample (if they are nominated as friends by other students). Since the students can be nominated by more than one other students, they can be alters of multiple ego-nets, and the model is a multiple-membership multiple-classification (MMMC) model.

$$y_i = \beta_0 + u_{\text{school(i)}} + \sum_{j \in \text{group(i)}} w_{i,j} u_j + e_i \tag{4.6}$$

where y_i is student i's test score, β_0 is the grand mean test score, $u_{\text{school(i)}}$ is the random effect of the school student i belongs to, $u_{\text{school(i)}} \sim \mathcal{N}(0, \sigma^2_{\text{sch}})$. The term $\sum_{j \in \text{group(i)}} w_{i,j} u_j$ involves J random effects, where J is the total number of ego-nets defined. group(i) is the ego-nets that student i is an alter of, group$(i) \subset (1, \ldots, J)$. $w_{i,j}$ is the weight assigned to each ego-net of student i is an alter of, and the weights sum up to one for each student. For example, if student i is nominated as a friend by three other students in the sample, $w_{i,j_1} = w_{i,j_2} = w_{i,j_3} = \frac{1}{3}$. u_j is the ego-net random effect, $u_j \sim \mathcal{N}(0, \sigma^2_{\text{ego}})$. e_i is the residual term, $e_i \sim \mathcal{N}(0, \sigma^2_{\mathrm{e}})$.

Fitting the model To fit the MMMC model, a weight matrix whose rows are individuals and columns are groups has to be specified. In our example, ego-nets are treated as groups. The weight matrix can be simply obtained by eliminating all-zero rows of the **D** matrix, transposing it and dividing its elements by the sums of rows. Note, since there are students that do not belong to any ego-nets, the weights associated these students are all zeros.

```
> D.row.sum <- rowSums(D)
> W         <- D[rowSums(D)!=0,]
> W         <- t(W)
> Weight    <- W/rowSums(W)
>
> Weight[is.na(Weight)] <- 0
```

Then we specify two design matrices required for the estimation of models with random effects directly. To do that, we need to generate all components required for estimation by assigning each individual a fake ego-net level and storing them in the an object (lmod).

```
> fake.ego.id <- rep(1:N.ego, length.out=N)
> mmmc        <- as.data.frame(cbind(mmmc, fake.ego.id))
> lmod        <- lFormula(y~(1|fake.ego.id)+(1|sch.id), data=mmmc)
```

In this process, a matrix named `Zt` and a list named `Ztlist` that both indicate memberships of individuals are generated. Since the ego-net random effect term is specified first, the first 651 (the number of ego-net levels) out of 661 (the sum of ego-nets and school levels) rows of the `Zt` matrix and the first (out of two) element of the `Ztlist` need to replaced with the right weight matrix.

```
lmod$reTrms$Zt[1:N.ego,] <- Matrix(t(Weight))
lmod$reTrms$Ztlist[[1]]  <- Matrix(t(Weight))
```

The next step is to let the program estimate the model using correct weight matrix.

```
> devfun       <- do.call(mkLmerDevfun, lmod)
> opt          <- optimizeLmer(devfun)
> MMMC.example <- mkMerMod(environment(devfun), opt, lmod$reTrms,
+                          fr = lmod$fr)
```

The results can be extracted with the `summary` function. Due to space constraint, only the random effect estimates are shown below. The estimates are very close identical to the true values (the *School+Network* model reported in Table 3 of Tranmer et al. (2014)).

```
> summary(MMMC.example)
Random effects:
 Groups      Name        Variance Std.Dev.
 fake.ego.id (Intercept) 0.2142   0.4629
 sch.id      (Intercept) 0.1135   0.3369
 Residual                0.8303   0.9112
```

The parameter estimates indicate that ego nets explain about 18% of the total variation (0.2142/[0.2142+0.1135+0.8303]), and schools explain about 10% of the total variation (0.1135/[0.2142+0.1135+0.8303]).

CHAPTER 5

Conclusion Remarks

Data with the cross-classified structure are ubiquitous in social and behavioral sciences. A canonical example of the cross-classified data in education is that students are cross-classified by schools they attend and neighborhood they stay in. Item response data also possess the cross-classified data structure, since item responses are cross-classified by persons and items. The cross-classified multilevel model (CCMM; Rasbash & Goldstein, 1994; Goldstein, 1994) has been proposed to model the cross-classified data. However, the applications of the CCMM are relatively scant and are mostly in the field of education, which can be partly attributed to the lack of computer software.

In the present study, we present the connections between the CCMM and a few measurement models and methods that are widely applied in psychology, including the random effect item response theory (IRT) model (Van den Noortgate et al., 2003), the model for rater effect, the multitrait-multimethod (MTMM) model (Campbell & Fiske, 1959), and the generalizability theory (G-theory) model (Shavelson & Webb, 1991). We also introduce several relatively little-known applications of the CCMM, including its applications to the meta analysis, social network analysis, and the study of immigrant students.

In addition, to address the issue of computer software, we introduce a flexible and efficient `R` package **PLmixed** that allows users to easily fit the CCMM. We show how the above-mentioned psychometric models, which can be viewed as CCMMs, applications of the CCMM, and the multiple membership multiple classification (MMMC) model can be fitted with **PLmixed**, with some aid of the `R` packages **lme4** and **blme**.

In sum, as the cross-classified data structure is prevalent in social and behavioral research, the CCMM would have many more applications with the support of computer software such as the `R` package **PLmixed**.

BIBLIOGRAPHY

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality*, *72*, 845–876.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, *1*, 103–124.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Sage Publications, Inc.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*, 1190–1208.

Cafri, G., Hedeker, D., & Aarons, G. A. (2015). An introduction and integration of cross-classified, multiple membership, and dynamic group random-effects models. *Psychological Methods*, *20*, 407.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Charlton, C., Rasbash, J., Browne, W. J., Healy, M., & Cameron, B. (2020). MLwiN Version 3.05. *Centre for Multilevel Modelling, University of Bristol*.

De Nooy, W. (2011). Networks of action and events over time. A multilevel discrete-time event history model for longitudinal network data. *Social Networks*, *33*, 31–40.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*, 1138–1151.

Doran, H., Bates, D., Bliese, P., Dowling, M., et al. (2007). Estimating the multilevel rasch model: With the lme4 package. *Journal of Statistical software*, *20*(2), 1–18.

Dorie, V., & Dorie, M. V. (2015). Package 'blme'. *Bayesian Linear Mixed-Effects Models. URL https://CRAN. R-project. org/package= blme.*

Dunn, E. C., Richmond, T. K., Milliren, C. E., & Subramanian, S. (2015). Using cross-classified multilevel models to disentangle school and neighborhood effects: an example focusing on smoking behaviors among adolescents in the United States. *Health & Place*, *31*, 224–232.

Fernández-Castilla, B., Maes, M., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2019). A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis. *Behavior Research Methods*, *51*, 1286–1304.

Fielding, A., & Goldstein, H. (2006). Cross-classified and multiple membership structures in multilevel models: An introduction and review.

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research*, *22*, 364–375.

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*, 941–954.

Guo, S., & Bollen, K. A. (2013). Research using longitudinal ratings collected by multiple raters: One methodological problem and approaches to its solution. *Social Work Research*, *37*, 85–98.

Hill, P. W., & Goldstein, H. (1998). Multilevel modeling of educational data with cross-classification and missing identification for units. *Journal of Educational and Behavioral Statistics*, *23*(2), 117–128.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2003). A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *166*, 279–300.

Jeon, M., Kaufman, C., & Rabe-Hesketh, S. (2019). Monte Carlo local likelihood approximation. *Biostatistics*, *20*, 164–179.

Jeon, M., & Rabe-Hesketh, S. (2012). Profile-likelihood approach for estimating generalized linear mixed models with factor structures. *Journal of Educational and Behavioral Statistics*, *37*, 518–542.

Jeon, M., & Rijmen, F. (2014). Recent developments in maximum likelihood estimation of mtmm models for categorical data. *Frontiers in Psychology*, *5*, 269.

Jeon, M., & Rockwood, N. (2017). *PLmixed: Estimate (generalized) linear mixed models with factor structures* [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=PLmixed`

Jeon, M., & Rockwood, N. (2018). PLmixed: An R Package for Generalized Linear Mixed Models With Factor Structures. *Applied Psychological Measurement*, *42*, 401–402.

Koster, J., Leckie, G., Miller, A., & Hames, R. (2015). Multilevel modeling analysis of dyadic network data with an application to ye'kwana food sharing. *American Journal of Physical Anthropology*, *157*, 507–512.

Lance, C. E., & Sloan, C. E. (1993). Relationships between overall and life facet satisfaction: A multitrait-multimethod (MTMM) study. *Social Indicators Research*, *30*, 1–15.

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*, 537–554.

Lei, X., Li, H., & Leroux, A. J. (2018). Does a teacher's classroom observation rating vary across multiple classrooms? *Educational Assessment, Evaluation and Accountability*, *30*, 27–46.

Levels, M., Dronkers, J., & Kraaykamp, G. (2008). Immigrant children's educational achievement in western countries: origin, destination, and community effects on mathematical performance. *American Sociological Review*, *73*, 835–853.

Maes, M., Van den Noortgate, W., Fustolo-Gunnink, S. F., Rassart, J., Luyckx, K., & Goossens, L. (2017). Loneliness in children and adolescents with chronic physical conditions: a meta-analysis. *Journal of Pediatric Psychology*, *42*, 622–635.

Murphy, D. L., & Beretvas, S. N. (2015). A comparison of teacher effectiveness measures calculated using three multilevel models for raters effects. *Applied Measurement in*

Education, *28*, 219–236.

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, *28*, 441–453.

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, 337–350.

Rasbash, J., Leckie, G., Pillinger, R., & Jenkins, J. (2010). Children's educational progress: partitioning family, school and area effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*, 657–682.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, DK: The Danish Institute of Educational Research.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer* (Vol. 1). Sage.

Tranmer, M., Steel, D., & Browne, W. J. (2014). Multiple-membership multiple-classification models for social network and group dependences. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *177*, 439–455.

Van den Noortgate, W., De Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics*, *28*, 369–386.

Vangeneugden, T., Laenen, A., Geys, H., Renard, D., & Molenberghs, G. (2005). Applying concepts of generalizability theory on clinical trial data to investigate sources of variation and their impact on reliability. *Biometrics*, *61*, 295–304.