Modern applications of cross-classified random effects models in social and behavioral research: Illustration with R package PLmixed

Sijia Huang[1]* and Minjeong Jeon[2]

[1]School of Education, Indiana University Bloomington, Bloomington, IN, United States, [2]School of Education and Information Studies, University of California, Los Angeles, Los Angeles, CA, United States

Cross-classified random effects models (CCREMs) have been developed for appropriately analyzing data with a cross-classified structure. Despite its flexibility and the prevalence of cross-classified data in social and behavioral research, CCREMs have been under-utilized in applied research. In this article, we present CCREMs as a general and flexible modeling framework, and present a wide range of existing models designed for different purposes as special instances of CCREMs. We also introduce several less well-known applications of CCREMs. The flexibility of CCREMs allows these models to be easily extended to address substantive questions. We use the free R package **PLmixed** to illustrate the estimation of these models, and show how the general language of the CCREM framework can be translated into specific modeling contexts.

KEYWORDS

cross-classified data, multilevel model, item response theory, multitrait-multimethod, network analysis, generalizability theory, PLmixed

1. Introduction

Data in social and behavioral studies often show multilevel structures (Bryk and Raudenbush, 1992). The multilevel structure may be purely hierarchical (or nested). In a purely hierarchical structure, lower-level units belong to one and only one higher-level unit. A canonical example of strictly nested data is found where students (at level 1) are nested within classrooms (at level 2), and classrooms are nested within schools (at level 3). A pure hierarchy as such may no longer hold when lower-level units belong to two or more types of higher-level units at the same level. This kind of situation yields a so-called cross-classified structure. An example of cross-classified data can be found where students (at level 1) are nested within schools and neighborhoods simultaneously (at level 2). Students who live in the same neighborhoods may not attend the same schools, and students who attend the same schools may be drawn from different neighborhoods; therefore the schools and neighborhoods are not nested but crossed. If lower-level units belong to two or more higher-level units of the same type, the data show a multiple membership structure, rather than crossed; for example, students attend more

than one classroom (multiple classroom membership) or more than one middle school (multiple school membership). We will focus on cross-classified structures, not multiple membership structures, in this paper.

Standard regression models designed for single-level data typically assume that units of analysis are independent of each other. This independence assumption is violated when the data show multilevel structures due to the clustering of units, i.e., dependence among the units within the same clusters. Hierarchical linear models (HLM; Bryk and Raudenbush, 1992), also called multilevel models, have been developed to address such lack of independence and account for dependence among units within the clusters where the units are nested. To accommodate more complex dependence structures due to cross-classified clustering in data, multilevel models have been extended to cross-classified random effects models (CCREMs; Raudenbush, 1993; Goldstein, 1994; Rasbash and Goldstein, 1994). In applied settings, researchers may fail to recognize the cross-classified structure of the data being analyzed, neglecting some or all crossed factors in data analysis. However, studies have shown that ignoring crossed factors can bias variance estimates and standard errors of the regression coefficients associated with predictors (Meyers and Beretvas, 2006; Luo and Kwok, 2009; Ye and Daniel, 2017).

To promote the use of CCREMs in applied research, we aim to introduce modern applications of CCREMs that may be less well-known in the literature. In particular, we will present several existing statistical/psychometric models as applications of CCREMs. Understanding these models as CCREMs can be beneficial; the flexibility of CCREMs allows researchers to extend these models to address important substantive issues by incorporating complex multilevel/cross-classified data structures, covariates, and other model structures. In addition, researchers can estimate these models and their extensions using existing software packages available for CCREMs. This means that researchers may not need to learn different software packages designed for specific models separately every time they need them.

We will illustrate the estimation of the discussed CCREM applications with a freely available R package, **PLmixed** (Jeon and Rockwood, 2018). **PLmixed** is a flexible R package designed to estimate models with nested and/or crossed random effects. It can estimate standard CCREMs as well as extended CCREMs with factor structures (e.g., factor loadings). The latter type, cross-classified models combined with measurement models, may not be estimable with other widely-used R packages, such as **lavaan** (Rosseel, 2012) and **mirt** (Chalmers, 2012). Further, **PLmixed** extends the popular R package **lme4** (Bates et al., 2015), so researchers who are familiar with **lme4** syntax can use **PLmixed** with minimal learning efforts. For each of the applications, we provide the **PLmixed** syntax and an example dataset (in

the manuscript and in the supplement), so that readers can easily apply and utilize the discussed models in their own research.

The remainder of this paper is organized as follows. In Section 2, we give a context to establish the notation of CCREMs. In Section 3, we briefly review available software options for CCREM estimation, and provide an overview of the **PLmixed** package. In Section 4, we present several statistical/psychometric models as special cases of CCREMs, and then present a couple of other less well-known, but interesting applications of CCREMs. Finally, in Section 5, we conclude the paper with a summary and a discussion on limitations and future directions.

2. Model

2.1. Setting

To set up the notation, we suppose a researcher wants to study the effects of elementary schools and neighborhoods on students' math performance, where the math performance was measured based on standardized test scores (continuous) or pass/fail evaluations (binary). The researcher also wants to explore how some characteristics of schools (e.g., public vs private), neighborhoods (e.g., residential income deprivation scores), and students (e.g., student age and race/ethnicity) may be related to the student performance measure. In this case students, the level-1 units, are cross-classified by two different level-2 units, schools and neighborhoods.

2.2. General formulation

Let y_{ijk} denote the math performance of student i ($i = 1, \ldots, I$) 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 specify a CCREM for y_{ijk} as follows:

$$g(E(y_{ijk}|u_j, u_k)) = \mathbf{X}_{ijk}\boldsymbol{\beta} + u_j + u_k, \quad (1)$$

where $g(\cdot)$ is a link function that transforms the linear predictor on the right-hand side to the conditional expectation $E(y_{ijk}|u_j, u_k)$. Using a link function allows us to accommodate outcome variables from various distribution families. For example, if y_{ijk} is continuous test scores and can be assumed to follow a normal distribution, an identity link function is used. If y_{ijk} is binary (e.g., pass/fail), a logit or probit link function can be used 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 (1), $\mathbf{X}_{ijk}$ is a covariate matrix, each row of which corresponds to a student and each column of which corresponds to a covariate, and $\boldsymbol{\beta}$ is a vector of regression coefficients for $\mathbf{X}_{ijk}$. The elements of the first column

of $\mathbf{X}_{ijk}$ are 1 when the first element of $\boldsymbol{\beta}$ is the intercept. With no covariates and continuous math scores y_{ijk}, the intercept is the grand mean, representing the *average* math scores of *average* students who attend an *average* school and lives in an *average* neighborhood.

In Equation (1), 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 as $u_j \sim \mathcal{N}(0, \sigma^2_{\text{school}})$ and $u_k \sim \mathcal{N}(0, \sigma^2_{\text{neighborhood}})$. Additional technical details on CCREMs can be found in comprehensive review articles, such as Fielding and Goldstein (2006) and Cafri et al. (2015).

3. Software options

In this section, we provide a brief review of several existing software programs and packages for estimating CCREMs. These include HLM 7 (Raudenbush et al., 2011), MLwiN (Charlton et al., 2020), ʀ package **lme4** (Bates et al., 2015), and functions in SAS 9.4 (SAS Inc, 2014). Although not discussed here to conserve space, additional software options are also available, such as Mplus (Muthén and Muthén, 1998–2017), SPSS (IBM Corp, 2020), and Bayesian options such as WinBUGS (Lunn et al., 2000) and Stan (Carpenter et al., 2017).

Subsequently, we describe the ʀ package **PLmixed** (Jeon and Rockwood, 2018), which is selected for illustration in the present article. **PLmixed** can be downloaded from CRAN (https://cran.r-project.org/web/packages/PLmixed/).

3.1. Existing programs and packages

3.1.1. HLM 7

HLM 7 (Raudenbush et al., 2011) is a commonly-used software program for fitting multilevel models. HLM 7 can fit two- and three-level CCREMs, using maximum likelihood estimation through the expectation–maximization (EM) algorithm (Raudenbush, 1993). For discrete outcome variables, HLM 7 estimates the models using the higher-order Laplace approximation. However, HLM 7 has limited capabilities in estimating CCREMs with more than three levels.

3.1.1.1. MLwiN

MLwiN (Charlton et al., 2020) is another widely-used statistical software program for fitting various types of multilevel models. MLwiN can handle both continuous and discrete outcome variables. MLwiN also allows for inconstant level-1 variance (i.e., heteroscedasticity). MLwiN implements the iterative generalized least squares (IGLS) and Markov chain Monte Carlo (MCMC) estimation procedures (Browne, 2015). Estimating CCREMs with IGLS can be computationally intensive. Thus, splitting the data into separate groups is recommended to reduce the amount of storage needed. For example, in the school and neighborhood example, if some schools only have students from certain neighborhoods, and the students from these neighborhoods do not go to other schools, these schools and neighborhoods can be treated as separate groups. However, using this trick may not always be possible in all data situations. Using the MCMC option might not be an easy route either, as fundamental knowledge about the Bayesian method would be required for proper usage of the option.

3.1.1.2. ʀ package lme4

lme4 (Bates et al., 2015) is a powerful ʀ package for maximum likelihood estimation of linear and generalized linear mixed models. For non-continuous outcome variables, **lme4** utilizes the (penalized) iteratively reweighted least squares using the Laplace approximation (Doran et al., 2007). **lme4** is highly flexible, allowing for many nested and crossed random effects at multiple levels. **lme4** has also been used to estimate one-parameter logistic (1PL) item response theory (IRT) models (e.g., Doran et al., 2007; De Boeck et al., 2011), and their various extensions, such as IRTree models (e.g., De Boeck and Partchev, 2012). However, **lme4** cannot estimate models that involve factor loadings, such as two-parameter IRT models and factor analysis models. **blme** (Dorie and Dorie, 2015), a Bayesian version of **lme4**, offers additional flexibility, but estimating complex CCREMs with factor loading structures may still be challenging with this package.

3.1.1.3. SAS 9.4

SAS 9.4 (SAS Inc, 2014) is a general-purpose statistical software program and it offers powerful procedures, such as GLIMMIX and MIXED, for fitting mixed models. GLIMMIX subsumes MIXED since it fits generalized linear mixed models, of which linear mixed models are a special case. GLIMMIX allows for accommodating complex nested and crossed random effects structures. GLIMMIX employs the restricted pseudo-likelihood (RPL) estimation (Wolfinger and O'connell, 1993), a linearization-based method, as its default estimation method for models with random effects. The PRL estimation method may yield biased parameter estimates especially for binary data. GLIMMIX also offers two integral approximation-based estimation methods, the Laplace approximation and adaptive Gauss-Hermite quadrature, but the latter may be computationally inefficient for estimating complex CCREMs.

3.2. ʀ package PLmixed

PLmixed extends the **lme4** package to estimate extended generalized linear mixed models with factor structures. **PLmixed** implements a profile maximum likelihood estimation approach (Jeon and

Rabe-Hesketh, 2012), taking advantage of the **lme4** estimation and the `optim` function (Byrd et al., 1995). **PLmixed** follows **lme4**'s syntax structure, and therefore, minimum learning may be required to utilize **PLmixed** for users who are already familiar with **lme4**.

To illustrate what **PLmixed** can do, let us return to the example used in Section 2.1. Suppose the researcher has tracked the students at multiple time points, and is interested in how the effects of schools change over time. For simplicity, assume the students did not change their school membership and residential areas (neighborhoods) during this time. In this longitudinal measurement context, we can extend the CCREM specified in Equation (1) as follows:

$$g(E(y_{itjk}|u_{jt}, u_k)) = \mathbf{X}_{itjk}\boldsymbol{\beta} + \lambda_t u_{jt} + u_k, \qquad (2)$$

where y_{itjk} indicates the i-th student's performance measure at time t ($t = 1, \ldots, T$) in school j and neighborhood k. Here the new parameter λ_t is introduced to model the weight of the school random effect u_{jt} at time t. The weight parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_t, \ldots, \lambda_T)'$ tell us how the impacts of schools on the students' math performance change over time. For identification purpose, one of the weights is fixed at 1 as a reference point (e.g., $\lambda_1 = 1$). Since the weight parameter λ_t is multiplied by random effect u_{jt} and indicates the size/direction of the relationship between the outcome measures across time and the random effect, they can be seen as factor loadings as in the factor analysis framework. For example, λ_t is the factor loading of the random effect u_{jt} (or latent variable or factor) on the t-th outcome variable y_{itjk}. In IRT, factor loadings are also called *item slopes* or *item discrimination parameters*. These factor loadings are not estimable with **lme4**.

The main function of the **PLmixed** package is `PLmixed`. The basic arguments of `PLmixed` function are identical to **lme4**'s `glmer` function's main arguments: `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 response variable). To estimate factor loadings, three new arguments 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 section through several CCREM applications. Of note, Rockwood and Jeon (2019) demonstrated the applications of **PLmixed** in the context of complex measurement and growth models, providing additional technical details of the profile likelihood estimation. We refer interested readers to Rockwood and Jeon (2019) for additional information.

4. Applications

In this section, we introduce several applications of CCREMs. We will first present several measurement models and methods as special applications of CCREMs. We will then introduce a couple of interesting CCREM applications that may be less well-known in applied research. For each of the applications, we first give a brief summary of the discussed models or methods. We then describe how to fit the models using **PLmixed** with an example dataset and briefly discuss the estimated results.

4.1. Measurement models and methods

4.1.1. Random item-effect IRT models
4.1.1.1. Background
IRT models describe the relationship between categorical item responses and the latent variable(s) to be measured with the test. The one-parameter logistic (1PL) model, also known as the Rasch model (Rasch, 1960), can be written for binary response y_{ij} to item i given by person j as follows:

$$\text{logit}(P(y_{ij} = 1 \mid \theta_j)) = \beta_i + \theta_j, \qquad (3)$$

where θ_j indicates person j's level on the latent variable or construct of interest; respondents are typically assumed drawn from a population distribution, assuming $\theta_j \sim \mathcal{N}(0, \sigma^2)$. β_i is the item intercept parameter representing the easiness level of item i. The 1PL model can also be seen as a two-level multilevel model with item responses at level 1 and respondents at level 2, where θ_j indicates person random effects and β_i indicates item fixed effects. The 1PL model can therefore be estimated with multilevel modeling packages, such as the R package **lme4** (Bates et al., 2015).

The 1PL model is extended to the two-parameter logistic (2PL) model by additionally accommodating the item slopes, or factor loadings α_i:

$$\text{logit}(P(y_{ij} = 1 \mid \theta_j)) = \beta_i + \alpha_i \theta_j, \qquad (4)$$

The item slope parameters α_i are multiplied by the latent variable $\theta_j \sim \mathcal{N}(0, \sigma^2)$, similar to Equation (2). For identifiability, one of the item factor loadings are fixed at 1 (or $\sigma = 1$). The 2PL model can be seen as a two-level multilevel model with a factor structure. Thus, **PLmixed** can estimate this model, while **lme4** cannot.

Van den Noortgate et al. (2003) presented an extended 1PL IRT model that assumes both respondents and items are random samples from the respective population distributions, therefore considered as random effects.

This so-called, random item-effect IRT model can be written as follows:

$$\text{logit}(P(y_{ij} = 1 \mid \theta_j, \delta_i)) = \beta_0 + \theta_j + \delta_i, \qquad (5)$$

where $\theta_j \sim \mathcal{N}(0, \sigma^2)$ and $\delta_i \sim \mathcal{N}(0, \tau^2)$ represent the respondent and item random effects, respectively, and β_0 indicates the intercept. In the sense that the same set of items are given to respondents, item responses are cross-classified by respondents and items. When both respondents and items are considered random effects, as in Equation (5), the model can be seen as a CCREM with no covariates.

Under the CCREM framework, the random item-effect IRT model described above can be easily extended with person and item covariates. For example,

$$\text{logit}(P(y_{ij} = 1 \mid \theta_j, \delta_i)) = \beta_0 + \mathbf{x}_j^{\text{P}} \boldsymbol{\beta}^{\text{P}} + \mathbf{x}_i^{\text{I}} \boldsymbol{\beta}^{\text{I}} + \mathbf{x}_{ij}^{\text{P,I}} \boldsymbol{\beta}^{\text{P,I}} + \theta_j + \delta_i,$$
(6)

where $\boldsymbol{\beta}^{\text{P}}$ is a vector of regression coefficients for $\mathbf{x}_j^{\text{P}}$, a vector of person j's characteristics (e.g., gender and motivation level). $\boldsymbol{\beta}^{\text{I}}$ is a regression coefficient vector for $\mathbf{x}_i^{\text{I}}$, a vector of item i's features (e.g., item length and cognitive complexity). $\boldsymbol{\beta}^{\text{P,I}}$ is a vector of regression coefficients of $\mathbf{x}_{ij}^{\text{P,I}}$, which is a vector of person-by-item interactions (e.g., whether person i's background, such as gender, interacts with item i or item i's feature). In addition, the CCREM framework allows for further expansions of the random item-effect IRT model with more complex person and item clustering structures; for example, respondents and/items are nested or cross-classified with higher-level units (e.g., respondents are cross-classified with schools and neighborhoods, and items are cross-classified with two or more cognitive functions).

4.1.1.2. Illustration

To show how to fit a random item-effect IRT model with **PLmixed**, we use the simulated data `IRTsim` in the package. The installation of the package and the first six rows of the data are presented below.

```
> install.packages("PLmixed")
> library(PLmixed)
> head(IRTsim)
    sid school item sch_x stu_x y
1.1   1     1     1 -1.21     0 0
1.2   1     1     2 -1.21     0 0
1.3   1     1     3 -1.21     0 0
1.4   1     1     4 -1.21     0 0
1.5   1     1     5 -1.21     0 0
2.1   2     1     1 -1.21     1 0
```

The data contain 6 variables and 2,500 binary item responses from a cognitive test. The first three columns of the data are `sid`, `school`, and `item`, indicating student ID, school ID, and item ID, respectively. There are a total of 500 students from 26 school and 5 items. `sch_x` represents a continuous school-level covariate. `stu_x` is a binary student-level covariate. `y` indicates students' binary responses that takes the value of 1 if the answer is correct and 0 otherwise.

Model To illustrate, we specify a random item-effect IRT model with a cross-classified structure of respondents with

schools. We include student-level and school-level covariates to examine their impacts on the success probability.

$$\begin{aligned}\text{logit}(P(y_{ijk} = 1 \mid \theta_{jk}, \theta_k, \delta_i)) = \beta_0 + \beta_1 \times \text{sch\_x}_k + \beta_2 \\ \times \text{stu\_x}_{jk} + \delta_i + \theta_{jk} + \theta_k,\end{aligned}$$
(7)

where β_0 is the intercept, β_1 is the regression coefficient for the school-level covariate, β_2 is the regression coefficient for the student-level covariate. $\delta_i \sim \mathcal{N}(0, \tau^2)$ is the random item effect, where τ indicates how much the item easiness varies across the test items. $\theta_{jk} \sim \mathcal{N}(0, \sigma_{\text{stu}}^2)$ and $\theta_k \sim \mathcal{N}(0, \sigma_{\text{sch}}^2)$ represent the student and school random effects, respectively.

Fitting the model The model can be fitted with the below **PLmixed** syntax:

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

The argument `formula` follows Equation (7). The item response `y` is a function of three fixed-effect and three random-effect terms. The inclusion of the term `1` indicates that the intercept of the linear predictor β_0 is estimated. With `sch_x` and `stu_x`, the regression coefficients, β_1 and β_2, are estimated as well. `(1|item)`, `(1|sid:school)` and `(1|school)` are the three random-effect terms. The colon between `sid` and `school` is used to indicate that the students are nested within schools. Here, the argument `family = binomial` means that the binomial family with a logit link function is used. Other than the default logit link, a probit link function can also be used with the syntax `family = binomial(link="probit")`. All results are saved in an object named `IRT.example`.

After fitting the model, the estimated results are summarized with the `summary()` function.

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

AIC      BIC    logLik deviance df.resid
 2716.01  2750.95 -1352.00  2245.42      2494

Scaled residuals:
    Min      1Q  Median      3Q      Max
-2.5110 -0.8657  0.4374  0.7744  2.1867

Random effects:
 Groups      Name        Variance Std.Dev.
 sid:school (Intercept) 0.6320   0.7950
 school     (Intercept) 0.7529   0.8677
 item       (Intercept) 0.3767   0.6138
```

```
Number of obs: 2500, groups:  sid:school,
 500; school, 26; item, 5

Fixed effects:
             Beta     SE z value  Pr(>|z|)
(Intercept) 0.2695 0.3432  0.7853 4.323e-01
sch_x       0.6963 0.2062  3.3771 7.326e-04
stu_x       1.3874 0.1326 10.4652 1.248e-25

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

The first section of the summary echos the formula we specified, the data, and the link function. The second section presents model fit indices including the AIC, BIC, and a summary of the scaled residuals.

In the `Random effects` section, the estimates of variances and standard deviations of the three random effects are presented. The last line of this section lists the numbers of observations, students, schools and items. These estimates indicate that about 42.7% (= 0.753/ [0.632 + 0.753 + 0.377]), 35.8% (= 0.632/ [0.632 + 0.753 + 0.377]), and 21.4% (= 0.377/ [0.632 + 0.753 + 0.377]) of the variation in the data are explained by the differences between schools, between students, and between items, respectively.

In the `Fixed effects` section, the estimates of the fixed effects are presented. The intercept estimate is 0.2695, indicating

```
> head(RATERsim)
  field    rating assessor assessor_gender proposal researcher_gender
1    1     0.75     19            0             446            0
2    1    -0.31     20            0             579            0
3    1    -0.04     20            0             184            0
4    1    -0.40     20            0              71            0
5    1    -0.46     28            1             117            0
6    1     0.43     34            0             293            1
```

that the probability that the an *average* student (i.e., $\theta_{jk} = 0$) in the stu_x=0 group from an *average* school (i.e., $\theta_k = 0$) with sch_x=0 correctly answers an *average* item (i.e., $\delta_i = 0$) is about 0.577 (= 1/ [1+exp(–0.2695)]). The change in the logit that is associated with one unit increase in sch_x is 0.6963, and the corresponding standard error (SE) is 0.2062. The average difference between the stu_x= 0 group and stu_x= 1 group is 1.3874, with SE of 0.1326.

The last section of the result lists `lme4` optimizer, `optim` optimizer, the number of `optim` iterations, and the estimation time.

4.1.2. Rater effect models
4.1.2.1. Background

Social science studies often utilize assessment by raters. In this case, raters contribute to the variance in the assessment data, and this variance due to rater effects is not related to the variation of the ratees in their performance (Scullen et al., 2000). Since rating data are cross-classified by raters and ratees, CCREMs are a natural option to address rater effects. For example, CCREMs have been used to estimate rater effects in cross-sectional data (e.g., Jayasinghe et al., 2003; Murphy and Beretvas, 2015; Lei et al., 2018; Martinez et al., 2020; Chen et al., 2021) and in longitudinal data settings (e.g., Guo and Bollen, 2013). Under the CCREM framework, researchers can differentiate rater variance from other sources of variances, and incorporate covariates to answer useful research questions. For example, a large rater variance indicates that ratings vary to a large degree across raters, implying that rater training needs to focus on improving consistency across raters (by making the scoring rubrics tighter and stricter, for instance). With covariates, one can find out whether particular characteristics of raters, ratees, and rater-and-ratee combinations may be associated with systematically higher or lower ratings.

4.1.2.2. Illustration

To show how to fit a rater effect model as a CCREM, we first generate a dataset that mimics the empirical data analyzed by Jayasinghe et al. (2003). The simulated data include 2,401 evaluations of 1,580 assessors on 673 grant proposals. Each assessor and proposal belongs to one of the 28 fields of study, which can be further categorized into nine general discipline panels, such as Physics and

Mathematics. The first six rows of the simulated dataset are printed below.

The outcome variable is the rating (`rating`) of a proposal (`proposal`) provided by an assessor (`assessor`) in a field of study (`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 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 data are the gender of the assessors (`assessor_gender`) and the gender of the first author of the proposals (`researcher_gender`). For these gender variables, 0 indicates the assessor/author is a male, and 1 indicates the assessor/author is a female.

Model We specify a three-level CCREM where ratings (at level 1) are cross-classified by assessors (raters) and proposals (at level 2) both of which are nested within fields (at level 3). In addition to the random effects of assessors, proposals, and fields, we examine the effects of the genders of assessors, first authors of proposals, and their interactions. The rater effect CCREM can then be specified as follows:

$$y_{ijk}|u_{ik}, u_{jk}, u_k = \beta_0 + \beta_1 x_{ik}^{R} + \beta_2 x_{jk}^{P} + \beta_3 x_{ijk}^{R,P} + u_{ik} + u_{jk} + u_k + \varepsilon_{ijk}, \tag{8}$$

where y_{ijk} is the rating of proposal j in field k assessor i provides. x_{ik}^{R}, x_{jk}^{P} and $x_{ijk}^{R,P}$ are the assessor' gender (1: female; 0: male), first author's gender (1: female; 0: male), and the interaction between the two (1: the assessor and author are both females; 0: the assessor and author are of different genders or are both males), respectively. β_0 is the intercept and β_1 to β_3 are the regression coefficients for the three gender-related covariates. The regression coefficient β_3 of the interaction term can answer an interesting question that whether female assessors' rating on the proposals written by female authors are consistently higher or lower. The assessor, proposal and field random effects are assumed $u_{ik} \sim \mathcal{N}(0, \sigma_{assessor}^2)$, $u_{jk} \sim \mathcal{N}(0, \sigma_{proposal}^2)$, and $u_{ijk} \sim \mathcal{N}(0, \sigma_{field}^2)$.

Fitting the model The below PLmixed syntax can be used to fit the model specified in Equation (8).

```
> 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)
```

This syntax is similar in structure to the syntax shown in Section 4.1.1. Thus, we omit giving all specific details of the arguments. The argument `family = gaussian` is used since the response data are continuous in this example. If `family` is omitted, Gaussian family with an identity link function is applied by default.

The `summary()` function gives the summary of the estimated results. To save space, we print here only the 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

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, the estimates of variances are shown. Consistent with the results shown in Jayasinghe et al. (2003), assessors explain about 82% (= 0.8876/ [0.8876+0.1672+0.0254]) of the total variation in ratings, proposals explain about 15% (= 0.1672/ [0.8876+0.1672+0.0254]), and fields explain 3% (= 0.0254/ [0.8876+0.1672+0.0254]) of the variation in the data. The large between-assessor variance tells us that improving consistency in ratings across assessors would be a useful consideration to improve the proposal assessment system.

In the `Fixed effects` section, the estimates of regression coefficients and the associated standard errors are presented. The results suggest that female assessors on average provide lower ratings than male assessors, female authors on average receive higher ratings than male authors, and proposals written by female authors receive lower ratings from female assessors.

4.1.3. Multitrait-multimethod models

4.1.3.1. Background

The multitrait-multimethod (MTMM) (Campbell and Fiske, 1959) models have been applied broadly in applied social science research (e.g., Lance and Sloan, 1993; Biesanz and West, 2004; DeYoung, 2006). When multiple traits (e.g., the Big Five) are measured with multiple methods (e.g., self report or peer report), MTMM models are used to differentiate method effects from trait effects. MTMM analysis helps researchers evaluate convergent and discriminant validity of the main measures; the methods designed to measure the same traits should be highly correlated, while the traits should not correlate highly with the methods.

The flexible CCREM framework allows researchers to extend the classical MTMM model in a variety of ways, e.g. by adding covariates and dealing with multilevel/cross-classified data structures. These extended models can be easily estimated with a CCREM package of choice, such as **PLmixed**, without needing to develop a new estimation routine.

4.1.3.2. Illustration

To show how the MTMM model can be estimated as a CCREM, we simulate a dataset that reproduces the MTMM covariance matrix reported by DeYoung (2006). The simulated data include continuous scores on the Big Five obtained from subjects ($n = 500$) and three of their peers. Each subject has 20 scores, where five are self-reported and the other 15 are reported by three peers (five per peer). The five self-reported scores are measures 1–5, and the 15 peer-reported scores are measures 6–10. The first six row of the simulated data are printed below.

```
> head(MTMM.data)
    subject measure method trait peer score
```


FIGURE 1
Circles in the figure represent latent variables: E, Extraversion; A, Agreeableness; C, Conscientiousness; N, Neuroticism; O, Openness/Intellect; P, peer rater effect. Squares are observed ratings: measures 1–5 are self-reported scores and measures 6–10 are peer-reported. The 1 in the figure represents fixed loading (imposed for identification purposes). The ? represents a loading that needs to be estimated.

```
[1,]      1     1      1     1     0    4.46
[2,]      1     2      1     2     0    3.62
[3,]      1     3      1     3     0    4.30
[4,]      1     4      1     4     0    4.30
[5,]      1     5      1     5     0    3.12
[6,]      1     6      2     1     1    5.13
```

The variable `subject`, `measure`, `method`, `trait` and `peer` are indicators of subject, scale, 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 Figure 1 illustrates the MTMM model that we consider for this dataset. Mathematically, the model is specified as follows:

$$y_{itpj}|\theta_{tj}^T, \theta_P^M = \beta_i + \lambda_{it}^T \theta_{tj}^T + \lambda_{ip}^M \theta_P^M + \varepsilon_{itpj}, \qquad (9)$$

where y_{itpj} represents peer p's rating to measure i on trait t of subject j. β_i denotes the intercept of measure i, λ_{it}^T indicates measure i's loading on θ_{tj}^T for t-th trait, and λ_{ip}^M indicates measure i's loading on θ_P^M, the peer rater effect. The trait latent variables and peer/rater latent variables are assumed: $\boldsymbol{\theta}_j^T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{T})$ and $\theta_P^M \sim \mathcal{N}(0, \sigma_M^2)$. The residual term is assumed $\varepsilon_{itpj} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

Fitting the model To fit the above model with `PLmixed`, we first specify the factor loading `lambda` matrix according to the diagram (Figure 1). Rows of the `lambda` matrix correspond to the measures, and columns correspond to the latent variables. For identification purpose, we fix the first loading of each `lambda` matrix to 1. Note that as measures 6–10 are peer-reported, 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 follows:
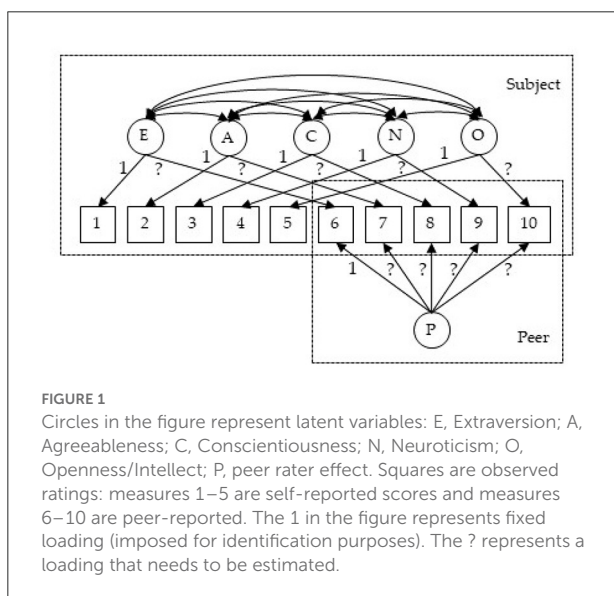
```
> MTMM.example <- PLmixed(score~0 + as.factor(measure)
+                        + (0+E+A+C+N+O|subject)
+                        + (0+rater|peer),
+                        data = as.data.frame(MTMM.data),
+                        lambda = list(lambda), load.var = "measure",
+                        factor = list(c("E","A","C","N","O","rater")))
```

The `0` is included to ensure no intercept is estimated. We use `as.factor(measures)` to estimate measure-specific intercepts. 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, the syntax would be specified as `(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 indicates the factor loading matrix we specified earlier. The `load.var` argument indicates the variable that defines the factor loading matrix. The `factor` shows the names of the latent variables.

Below shows the summary of the estimated results.

```
>summary(MTMM.example)
Profile-based Mixed Effect Model Fit With PLmixed Using lme4
Formula:  score~0 + as.factor(measure) + (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:  measure
      E   SE    A   SE    C   SE    N   SE    O   SE  rater   SE
1     1    .    .    .    .    .    .    .    .    .    .      .
2     .    .    1    .    .    .    .    .    .    .    .      .
3     .    .    .    .    1    .    .    .    .    .    .      .
4     .    .    .    .    .    .    1    .    .    .    .      .
5     .    .    .    .    .    .    .    .    1    .    .      .
6   0.91 0.04   .    .    .    .    .    .    .    .    1      .
7     .    . 0.99 0.07   .    .    .    .    .    . 0.91   0.06
8     .    .    .    . 0.83 0.04   .    .    .    . 0.83   0.04
9     .    .    .    .    .    . 1.84 0.14   .    . 1.54   0.12
10    .    .    .    .    .    .    .    . -2.4 0.18 1.16   0.10


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(measure)1    3.336 0.03687   90.48
as.factor(measure)2    4.064 0.02723  149.23
as.factor(measure)3    4.060 0.02803  144.86
as.factor(measure)4    2.572 0.03515   73.18
as.factor(measure)5    3.670 0.03252  112.87
as.factor(measure)6    3.670 0.03081  119.12

as.factor(measure)7    4.104 0.02438  168.31
as.factor(measure)8    4.185 0.02324  180.06
as.factor(measure)9    2.610 0.03065   85.17
as.factor(measure)10   3.734 0.02513  148.58


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

The first section of the results echos the formula, and the second section shows model fit indices. The `Lambda` section shows the estimates of factor loadings and the corresponding standard errors. The negative loading of measure 10 on the latent variable O indicates that the peer-reported scores are negatively associated with *Openness/Intellect*, suggesting that these peer-reported scores of measure 10 should be closely examined as they do not seem to properly measure the target latent variable *Openness/Intellect*. Following the factor loading section, the `Random effects` section lists the estimates of variances of all latent variables, and the correlations of the five traits. In the `Fixed effects` section, measure-specific intercepts are presented. In the last section, information about **lme4** optimizer is listed.

4.1.4. Generalizability theory
4.1.4.1. Background

Generalizability theory (G-theory) is a statistical framework for evaluating the generalizability (or reliability) of measurements (Shavelson and Webb, 1991). In G-theory, observed scores 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 levels, or if researchers would like to generalize beyond these observed levels, this facet is a *random* facet. In contrast, if all possible levels of a facet are included in the design, or if there is no need for generalization, this facet is treated as a *fixed* facet. G-theory includes two types of studies, the generalizability study (G study) and the decision study (D study). In a G study, variances associated with the facets are computed. In a D study, different generalizability coefficients are constructed, using the variance estimates computed from the G study. A generalizability coefficient is analogous to a

reliability coefficient, while it depends on which facets are considered random.

Vangeneugden et al. (2005) utilized a linear mixed model framework to estimate G-theory models, so that different kinds of generalizability/reliability coefficients can be derived from the estimates of variances. Choi (2013) and Choi and Wilson (2018) presented a generalized linear latent and mixed modeling approach (GLLAMM) (Skrondal and Rabe-Hesketh, 2004) to combine G-theory and IRT. Jiang (2018) and Jiang et al. (2020) showed how generalizability variance components from univariate and multivariate generalizability theory can be estimated within the linear mixed model framework with general-purpose R packages.

4.1.4.2. Illustration

To illustrate how a G-theory model can be estimated and extended as a CCREM, we analyze the `Brennan.3.2` dataset from the **gtheory** R package (Moore, 2016). 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). The first six rows of the data are printed below.

```
> 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 person×(rater:task) design, there are five sources of variation: persons, tasks, raters, person-task interactions and residuals. The data show a cross-classified structure where the scores are cross-classified by persons and raters, which are nested within tasks. We specify the G-theory model as a CCREM as follows:

$$y_{ijk}|u_{jk}, u_k, u_{ik} = \beta_0 + u_i + u_{jk} + u_k + u_{ik} + \varepsilon_{ijk}, \quad (10)$$

where y_{ijk} is person i's score on task k by rater j, $u_i \sim \mathcal{N}(0, \sigma_{\text{person}}^2)$ represents the person random effect, $u_{jk} \sim \mathcal{N}(0, \sigma_{\text{rater}}^2)$ is the rater random effect, $u_k \sim \mathcal{N}(0, \sigma_{\text{task}}^2)$ is the task random effect, $u_{ik} \sim \mathcal{N}(0, \sigma_{\text{person}\times\text{task}}^2)$ is the person-task interaction random effect, and $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_e^2)$ is the residual term.

Fitting the model The `PLmixed` syntax to fit this model is displayed below.

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

The summary of the estimated results is shown below; only the `Random effects` and `Fixed effects` sections are listed to conserve space.

```
> 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 the estimates of variances associated with multiple variance sources, we can derive desired generalizability/reliability coefficients. For example, if we want to generalize persons' scores over raters and tasks, a generalizability coefficient is constructed as:

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

The reliability coefficient above is fairly low (consistent with the result from the **gtheory** package). This means that it would be difficult to generalize subjects' scores over different raters and tasks, calling for further investigation on the assessment data.

4.2. Other applications

4.2.1. Social network analysis
4.2.1.1. Background
Social network analysis (SNA) examines how individuals are connected and how social connections influence individuals and their behaviors (Otte and Rousseau, 2002). A number of methods and techniques have been developed for social network analysis, such as stochastic actor oriented models (SAOM) (e.g., Snijders, 1996, 2017), exponential random graph models (ERGM) (e.g., Holland and Leinhardt, 1981; Frank and Strauss, 1986; Cranmer and Desmarais, 2011), and latent space models (e.g., Hoff et al., 2002; Sewell and Chen, 2015).

CCREMs have also been utilized to study social network connections and dependencies among individuals generated due to their social networks. For example, Tranmer et al. (2014) considered individuals' ego-nets (of size 2 and 3) as their classification structures and applied a CCREM to account for the dependencies due to the network structure. They found that ignoring network dependencies as such would bias the estimates of fixed and random effect parameters in the model. De Nooy (2011) analyzed reviews/interviews among 40 literary authors and critics in the Netherlands during 1970–1980. The network model adopted in this study can be viewed as a CCREM in the sense that the model viewed review/interviews were cross-classified by authors and critics. Koster et al. (2015) showed that the multilevel formulation of the Social Relations Model (SRM) (Snijders and Kenny, 1999) could be estimated as a cross-classified Poisson model. The SRM as a CCREM partitions the total variance in social network data into variance components contributed by multiple sources.

4.2.1.2. Illustration
To illustrate how a CCREM can be used for social network analysis, we analyze a dyadic network dataset used by Koster et al. (2015). The data were collected in a village of indigenous Ye'kwana horticulturalists with eight households in Venezuela. These eight households lead to 28 dyads. The first six rows of the data are printed below:

```
> network <- network[order(network$dyad),]
> head(network)
   giver receiver dyad relationship sharing association distance kinship
1      1        2    2           12       1           1     0.234      162  0.0156
8      2        1    1           12       8           0     0.233      162  0.0160
2      1        3    1           13       2           1     0.144      286  0.0000
15     3        1    1           13      15           5     0.144      286  0.0000
3      1        4    1           14       3           0     0.120      327  0.0000
22     4        1    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, meaning that the values may not be the same for household A-B and B-A pairs. Other covariates are the distance between households (`distance`), the genetic relatedness between households (`kinship`), and an association index which is a measure of interactions between each pair of households (`association`).

Model The number of meals y_{ij} that household i gives to household j is assumed to follow a Poisson distribution with mean μ_{ij}, i.e., $y_{ij} \sim \text{Pois}(\mu_{ij})$. For simplicity, we consider no covariates and specify a Poisson CCREM as follows:

$$\log(\mu_{ij}) = \beta_0 + g_i + r_j + u_{|ij|} + \varepsilon_{ij}, \tag{11}$$

where β_0 is the intercept, g_i and r_j represent the giver and receiver random effects, respectively. The two random effects are assumed to be independent with $g_i \sim \mathcal{N}(0, \sigma^2_{\text{giver}})$ and $r_j \sim \mathcal{N}(0, \sigma^2_{\text{receiver}})$. We further decompose the relationship random effects into two parts: $u_{|ij|}$ represents the symmetric (undirected) relationship random effect, with $u_{|ij|} \sim \mathcal{N}(0, \sigma^2_{\text{dyad}})$. The $|\cdot|$ is used in the subscript to indicate that this effect is symmetric, meaning that it takes the same value within each dyad. ε_{ij} is the asymmetric (directed) relationship random effect, with $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\text{relation}})$. This effect is asymmetric, meaning that the value of ε_{ij} can be different for each observation within a dyad. Covariates at the giver, receiver and relationship levels can be also incorporated into the model.

Fitting the model The network model described above can be fit with the below `PLmixed` syntax.

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

We show below the `Random effects` and `Fixed effects` sections of the estimated results.

```
> 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 the variance estimates of the random effects, we can quantify the contribution of each variation source to the total variance. For example, the amount of total variation attributed to the givers is 0.549 (= 1.8853/[0.2401 + 1.1990 + 0.1060 + 1.8853]); that is, givers explain 54.9% of the total variation in the data. In contrast, receivers explain only 3.1% of the variance (0.031 = 0.1060/[0.2401 + 1.1990 + 0.1060 + 1.8853]).

4.2.2. Effects of country of origin and destination

4.2.2.1. Background

Another interesting application of CCREMs is found in immigrant studies that examine impacts of countries of origin and destination on various outcomes of immigrants. Examples of such studies include Levels et al. (2008), Kislev (2017), and Dronkers et al. (2014).

4.2.2.2. Illustration

For illustration, we simulate data that mimic the PISA data analyzed by Levels et al. (2008). The simulated data include the math scores of 7,403 immigrant students who were born in 35 different countries and took the PISA math test in 13 different countries. Below we print out the first six rows of the dataset.

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

Model In this example, students are cross-classified by countries of origin and destination. The communities indicate the origin-by-destination combinations. To differentiate the three types of contextual effects on students' math scores, we specify the following CCREM model:

$$y_{ijk}|u_j, u_k, u_{jk} = \beta_0 + u_j + u_k + u_{jk} + \varepsilon_{ijk}, \tag{12}$$

where y_{ijk} 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 the origin and destination countries. β_0 is the intercept, $u_j \sim \mathcal{N}(0, \sigma^2_{\text{origin}})$ and $u_k \sim \mathcal{N}(0, \sigma^2_{\text{destination}})$ are the origin country and destination country random effects, respectively, and $u_{jk} \sim \mathcal{N}(0, \sigma^2_{\text{community}})$ is the community random effects, as specified as the cell effects (cells created by origin and destination country classifications). $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$ is the residual term.

Fitting the model The above model can be estimated with the `PLmixed` syntax below:

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

Elements in the `formula` argument correspond to the terms in Equation (12). The `1` represents the intercept, and `(1|origin)`, `(1|destination)` and `(1|origin:destination)` are the three random-effect terms.

Below we display the `Random effects` and `Fixed effects` sections of the estimated results.

```
> summary(PISA.example)
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
```

The `Random effects` section lists the variance estimates of random effects. Here `origin:destination` indicates the community random effects specified as the cell effects. The results suggest that the immigrant community, the country of origin, and the country of destination explain 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% ($= 681.4/[511.2 + 803.2 + 681.4 + 7459.7]$) of the total variation in the math scores, respectively. It is interesting that the country of origin still has influence immigrant students' math scores, and the students' immigrant community shows as large influence as the destination county on immigrant students' math outcomes.

5. Concluding remarks

In this article, we aim to introduce some modern applications of CCREMs for social science studies. We presented some existing models and methods widely used in applied research as applications of CCREMs. These applications and their connections to CCREMs might have been shown in the literature by different authors in separate studies. However, to our best knowledge, these different applications have not been reviewed, discussed, and integrated together in a single paper.

We want to stress again the usefulness of understanding the connections between different models that serve for seemingly different purposes within the CCREM framework. As discussed earlier, those models can easily be extended to accommodate complex clustering and covariate structures in the data for any units of analysis. Further, the models presented in this paper can be integrated when desirable under the unifying framework that CCREM offers. For example, the rater effect model can be combined with the random item-effect IRT model, and/or with the multitrait-multimethod model, and/or the G-theory model. As another example, the model presented for a study of country of origin and destination may be extended with a measurement model such as the two-parameter IRT model. Without having these perspectives, researchers may need to develop their own estimation routines every time there is a need for such models, which can be inefficient, time consuming, and unnecessary.

It may be worth commenting that the flexibility and generality of the CCREM framework come with a cost in the sense that researchers need to understand the CCREM framework, and translate the general language of the CCREM framework into the specific modeling contexts of choice, which may not be an easy task. The current article is written to help remediate such a challenge by providing a practical guide to the CCREM and its applications.

Lastly, we chose the `R` package **PLmixed** to illustrate the estimation of the discussed models. **PLmixed** is a convenient choice as it is freely available and it can estimate all of the models presented in this paper and many of their combinations.

Of note, when no factor structures are involved, the models presented in this paper can also be estimated with the `R` package **lme4**. But when factor loadings are included, **lme4** can no longer be used. In addition, an important future line of work is to compare **PLmixed** with other packages specialized for specific models as well as other general-purpose software programs in terms of the model specification and estimated results. Further, it would also be useful to conduct rigorous simulation studies to compare the performance accuracy of different software packages for estimating different CCREM applications introduced in this paper.

Author's note

SH is currently an Assistant Professor at Indiana University Bloomington. MJ is an Associate Professor at the University of California, Los Angeles.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SH and MJ drafted the manuscript. Both authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Opinions are those of the authors alone and do not necessarily reflect those of the authors' home institutions.

References

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Biesanz, J. C., and 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. *J. Pers.* 72, 845–876. doi: 10.1111/j.0022-3506.2004.00282.x

B. M., Corp (2020). *IBM SPSS Statistics for Windows, Version 27.0.* Armonk, NY: IBM Corp.

Browne, W. J. (2015). *MCMC Estimation in MLwiN.* Centre for Multilevel Modelling, University of Bristol.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage Publications, Inc.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* 16, 1190–1208. doi: 10.1137/0916069

Cafri, G., Hedeker, D., and Aarons, G. A. (2015). An introduction and integration of cross-classified, multiple membership, and dynamic group random-effects models. *Psychol. Methods* 20, 407–421. doi: 10.1037/met0000043

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105. doi: 10.1037/h0046016

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: a probabilistic programming language. *J. Stat. Softw.* 76, 1–32. doi: 10.18637/jss.v076.i01

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *J. Stat. Softw.* 48, 1–29. doi: 10.18637/jss.v048.i06

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

Chen, J. M., Norman, J. B., and Nam, Y. (2021). Broadening the stimulus Set: introducing the American multiracial faces database. *Behav. Res. Methods* 53, 371–389. doi: 10.3758/s13428-020-01447-8

Choi, J. (2013). *Advances in combining generalizability theory and item response theory* (Ph.D. thesis). University of California, Berkeley: Berkeley, CA.

Choi, J., and Wilson, M. R. (2018). Modeling rater effects using a combination of generalizability theory and IRT. *Psychol. Test Assess. Model.* 60, 53–80. Available online at: https://www.psychologie-aktuell.com/fileadmin/download/ptam/1-2018_20180323/4_PTAM_Choi___Wilson__2018-03-10__2032.pdf

Cranmer, S. J., and Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Polit. Anal.* 19, 66–86. doi: 10.1093/pan/mpq037

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., et al. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *J. Stat. Softw.* 39, 1–28. doi: 10.18637/jss.v039.i12

De Boeck, P., and Partchev, I. (2012). IRTrees: tree-based item response models of the GLMM family. *J. Stat. Softw.* 48, 1–28. doi: 10.18637/jss.v048.c01

De Nooy, W. (2011). Networks of action and events over time. A multilevel discrete-time event history model for longitudinal network data. *Soc. Netw.* 33, 31–40. doi: 10.1016/j.socnet.2010.09.003

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *J. Pers. Soc. Psychol.* 91, 1138–1151. doi: 10.1037/0022-3514.91.6.1138

Doran, H., Bates, D., Bliese, P., and Dowling, M. (2007). Estimating the multilevel rasch model: with the lme4 package. *J. Stat. Softwaren* 20, 1–18. doi: 10.18637/jss.v020.i02

Dorie, V., and Dorie, M. V. (2015). *Package 'Blme'. Bayesian Linear Mixed-Effects Models.* Available online at: https://CRAN.R-project.org/package=blme

Dronkers, J., Levels, M., and de Heus, M. (2014). Migrant pupils' scientific performance: the influence of educational system features of origin and destination countries. *Largescale Assess. Educ.* 2, 1–28. doi: 10.1186/2196-0739-2-3

Fielding, A., and Goldstein, H. (2006). *Cross-Classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review. Research Report RR791.* London: Department for Education and Skills. Available online at: http://www.education.gov.uk/publications/eOrderingDownload/RR791

Frank, O., and Strauss, D. (1986). Markov graphs. *J. Am. Stat. Assoc.* 81, 832–842. doi: 10.1080/01621459.1986.10478342

Goldstein, H. (1994). Multilevel cross-classified models. *Sociol. Methods Res.* 22, 364–375. doi: 10.1177/0049124194022003005

Guo, S., and Bollen, K. A. (2013). Research using longitudinal ratings collected by multiple raters: one methodological problem and approaches to its solution. *Soc. Work. Res.* 37, 85–98. doi: 10.1093/swr/svs029

Hoff, P., Raftery, A., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *J. Am. Stat. Assoc.* 97, 1090–1098. doi: 10.1198/016214502388618906

Holland, P. W., and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *J. Am. Stat. Assoc.* 76, 33–50. doi: 10.1080/01621459.1981.10477598

Jayasinghe, U. W., Marsh, H. W., and 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. *J. R. Stat. Soc. A* 166, 279–300. doi: 10.1111/1467-985X.00278

Jeon, M., and Rabe-Hesketh, S. (2012). Profile-likelihood approach for estimating generalized linear mixed models with factor structures. *J. Educ. Behav. Stat.* 37, 518–542. doi: 10.3102/1076998611417628

Jeon, M., and Rockwood, N. (2018). PLmixed: an R package for generalized linear mixed models with factor structures. *Appl. Psychol. Meas.* 42, 401–402. doi: 10.1177/0146621617748326

Jiang, Z. (2018). Using the linear mixed-effect model framework to estimate generalizability variance components in r: a lme4 package application. *Methodol. Eur. J. Res. Methods Behav. Soc. Sci.* 14, 133. doi: 10.1027/1614-2241/a000149

Jiang, Z., Raymond, M., Shi, D., and DiStefano, C. (2020). Using a linear mixed-effect model framework to estimate multivariate generalizability theory parameters in R. *Behav. Res. Methods* 52, 2383–2393. doi: 10.3758/s13428-020-01399-z

Kislev, E. (2017). Deciphering the 'ethnic penalty' of immigrants in Western Europe: a cross-classified multilevel analysis. *Soc. Indic. Res.* 134, 725–745. doi: 10.1007/s11205-016-1451-x

Koster, J., Leckie, G., Miller, A., and Hames, R. (2015). Multilevel modeling analysis of dyadic network data with an application to Ye'kwana food sharing. *Am. J. Phys. Anthropol.* 157, 507–512. doi: 10.1002/ajpa.22721

Lance, C. E., and Sloan, C. E. (1993). Relationships between overall and life facet satisfaction: a multitrait-multimethod (MTMM) study. *Soc. Indic. Res.* 30, 1–15. doi: 10.1007/BF01080329

Lei, X., Li, H., and Leroux, A. J. (2018). Does a teacher's classroom observation rating vary across multiple classrooms? *Educ. Assess. Evaluat. Accountabil.* 30, 27–46. doi: 10.1007/s11092-017-9269-x

Levels, M., Dronkers, J., and Kraaykamp, G. (2008). Immigrant children's educational achievement in western countries: origin, destination, and community effects on mathematical performance. *Am. Sociol. Rev.* 73, 835–853. doi: 10.1177/000312240807300507

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* 10, 325–337. doi: 10.1023/A:1008929526011

Luo, W., and Kwok, O.-,m. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behav. Res.* 44, 182–212. doi: 10.1080/00273170902794214

Martinez, J. E., Funk, F., and Todorov, A. (2020). Quantifying idiosyncratic and shared contributions to judgment. *Behav. Res. Methods* 52, 1428–1444. doi: 10.3758/s13428-019-01323-0

Meyers, J. L., and Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behav. Res.* 41, 473–497. doi: 10.1207/s15327906mbr4104_3

Moore, C. (2016). *Gtheory: Apply Generalizability Theory With R. R Package Version 0.1.2.* Available online at: https://CRAN.R-project.org/package=gtheory

Murphy, D. L., and Beretvas, S. N. (2015). A comparison of teacher effectiveness measures calculated using three multilevel models for raters effects. *Appl. Meas. Educ.* 28, 219–236. doi: 10.1080/08957347.2015.1042158

Muthén, B. and Muthén, L. (1998–2017). *Mplus User's Guide, 8th Edn.* Los Angeles, CA: Muthén & Muthén.

Otte, E., and Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *J. Inform. Sci.* 28, 441–453. doi: 10.1177/016555150202800601

Rasbash, J., and Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *J. Educ. Behav. Stat.* 19, 337–350. doi: 10.3102/10769986019004337

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, DK: The Danish Institute of Educational Research.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *J. Educ. Stat.* 18, 321–349. doi: 10.3102/10769986018004321

Raudenbush, S. W., Bryk, A., Cheong, Y., Congdon, R., and Du Toit, M. (2011). *HLM 7 [Computer software]*. Lincolnwood, IL: Scientific Software International.

Rockwood, N. J., and Jeon, M. (2019). Estimating complex measurement and growth models using the R Package PLmixed. *Multivariate Behav. Res.* 54, 288–306. doi: 10.1080/00273171.2018.1516541

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i,02

SAS Inc. (2014). *Base SAS®9.4 Procedures Guide: Statistical Procedures*. Cary, SAS Institute Inc.

Scullen, S. E., Mount, M. K., and Goff, M. (2000). Understanding the latent structure of job performance ratings. *J. Appl. Psychol.* 85, 956. doi: 10.1037/0021-9010.85.6.956

Sewell, D. K., and Chen, Y. (2015). Latent space models for dynamic networks. *J. Am. Stat. Assoc.* 110, 1646–1657. doi: 10.1080/01621459.2014.988214

Shavelson, R. J., and Webb, N. M. (1991). *Generalizability Theory: A Primer*, Vol. 1. Newbury Park, CA: Sage Publications, Inc.

Skrondal, A., and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.

Snijders, T. A. (1996). Stochastic actor-oriented models for network change. *J. Math. Sociol.* 21, 149–172. doi: 10.1080/0022250X.1996.9990178

Snijders, T. A. (2017). Stochastic actor-oriented models for network dynamics. *Ann. Rev. Stat. Appl.* 4, 343–363. doi: 10.1146/annurev-statistics-060116-054035

Snijders, T. A., and Kenny, D. A. (1999). The social relations model for family data: a multilevel approach. *Pers. Relatsh.* 6, 471–486. doi: 10.1111/j.1475-6811.1999.tb00204.x

Tranmer, M., Steel, D., and Browne, W. J. (2014). Multiple-membership multiple-classification models for social network and group dependences. *J. R. Stat. Soc. A* 177, 439–455. doi: 10.1111/rssa.12021

Van den Noortgate, W., De Boeck, P., and Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *J. Educ. Behav. Stat.* 28, 369–386. doi: 10.3102/10769986028004369

Vangeneugden, T., Laenen, A., Geys, H., Renard, D., and 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. doi: 10.1111/j.0006-341X.2005.031040.x

Wolfinger, R., and O'connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *J. Stat. Comput. Simul.* 48, 233–243. doi: 10.1080/00949659308811554

Ye, F., and Daniel, L. (2017). The impact of inappropriate modeling of cross-classified data structures on random-slope models. *J. Modern Appl. Stat. Methods* 16, 458–484. doi: 10.22237/jmasm/1509495900