Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



British Journal of Mathematical and Statistical Psychology (2007), 60, 351–365
© 2007 The British Psychological Society



The British Psychological Society
www.bpsjournals.co.uk

# Conditional mixed models with crossed random effects

Fabian S. Tibaldi[1]*, Geert Verbeke[2], Geert Molenberghs[3], Didier Renard[1], Wim Van den Noortgate[2] and Paul de Boeck[2]

[1]Eli Lilly and Co., Belgium
[2]Katholieke Universiteit Leuven, Belgium
[3]Hasselt University, Belgium

The analysis of continuous hierarchical data such as repeated measures or data from meta-analyses can be carried out by means of the linear mixed-effects model. However, in some situations this model, in its standard form, does pose computational problems. For example, when dealing with crossed random-effects models, the estimation of the variance components becomes a non-trivial task if only one observation is available for each cross-classified level. Pseudolikelihood ideas have been used in the context of binary data with standard generalized linear multilevel models. However, even in this case the problem of the estimation of the variance remains non-trivial. In this paper, we first propose a method to fit a crossed random-effects model with two levels and continuous outcomes, borrowing ideas from conditional linear mixed-effects model theory. We also propose a crossed random-effects model for binary data combining ideas of conditional logistic regression with pseudolikelihood estimation. We apply this method to a case study with data coming from the field of psychometrics and study a series of items (responses) crossed with participants. A simulation study assesses the operational characteristics of the method.

## 1. Introduction

Models where population units are hierarchically structured have been studied extensively over several decades. However, in many cases, units at the same level of a hierarchy are simultaneously classified by more than one factor. As an example, let us consider the case where school pupils are classified by the school they attend, as well as by the neighbourhood in which they live. Since schools usually attract pupils from several neighbourhoods and children from the same neighbourhood usually attend several schools, these two factors are crossed, i.e. one factor is not nested within the

* *Correspondence should be addressed to Fabian S. Tibaldi, Eli Lilly and Co., Rue Granbonpre 11, Mont-Saint-Guibert 1348, Belgium (e-mail: tibaldifa@lilly.com).*

DOI:10.1348/000711006X110562

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



levels of the other. While crossed random-effects models extend classical hierarchical multilevel models, they can be fitted using procedures designed for purely hierarchical or multilevel structures (Bryk & Raudenbush, 1992; Dempster, Rubin, & Tsutakawa, 1981; Fahrmeir & Tutz, 2001; Laird & Ware, 1982; Verbeke & Molenberghs, 2000). Of course, they are technically and computationally more demanding and can become prohibitive since the data cannot be grouped into independent blocks. In the context of continuous outcomes, we apply conditional linear mixed model ideas (Verbeke, Spiessens, & Lesaffre, 2001) to conveniently estimate parameters, as well as precision. The major advantage of the approach is that, by appropriate conditioning, the original model maps into two hierarchical ones, for which conventional and hence computationally efficient and fast techniques can be used. The price for this is that no cross-sectional effects can be estimated, so the method is of use when within-cluster effects and in-variance components are of interest.

The aforementioned approach assumes the response variables (targets) to be continuous. This is justifiable when targets are sums of a sufficient number of dichotomous items. However, when an analysis at item level is needed, treating them as continuous is no longer tenable. Therefore, we also developed a solution to this problem and we present a method based on a combination of conditional logistic regression and pseudolikelihood methods for the estimation of the model parameters. While the solution is elegant and can handle clusters of arbitrary size, the methodology necessarily is more complicated than in the continuous case and therefore we introduce the method for the special case of two outcomes, because then well-understood ideas from matched pairs can be used. It is envisaged that extension to the general case will not pose insurmountable problems. Of course, when such a generalization is considered, it would be imperative to study such issues as convergence.

This paper is organized as follows. In Section 2, the case study which illustrates our method is introduced. In Section 3, the cross-classification multilevel models in a psychometric context are described. The methodology is fully described in Section 4 for the continuous case and in Section 5 for the binary situation. The case is analysed in Sections 6 and 7. A discussion follows in Section 8.

## 2. Case study

The Flemish community in Belgium issued a set of attainment targets that specify the basic competences that are expected from pupils leaving primary education. De Boeck, Daems, Meulders, and Rymenams (1997) explored the assessment of the attainment targets of reading comprehension in Dutch. These attainment targets can be characterized by the text type and by the level of processing. In the example, we use the data of one of the tests that was developed by De Boeck *et al.* and studied by Van den Noortgate, De Boeck, and Meulders (2005). These data have been analysed previously by Janssen, Tuerlinckx, Meulders, and De Boeck (2000). The data consist of the responses of a group of 539 pupils from 15 schools who answered 57 items assumed to measure nine attainment targets. In Table 1, the nine attainment targets are described by the type of text and by the level of processing. In addition, we indicate the number of items that were used to measure each one of the targets.

The advantage of this set of data is that it can be used to illustrate the various methodologies in an integrated fashion. In the first part of this analysis, we will use as targets as the sum of all positive answers within a category and assume them to be continuous, in line with general practice in the field (Van den Noortgate *et al.*, 2005), first

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



**Table 1.** Text type, level of processing and number of items for the attainment targets of the text

| Target | Text type | Level of processing | No. of items |
|---|---|---|---|
| 1 | Instructions | Retrieving | 4 |
| 2 | Articles in magazine | Retrieving | 6 |
| 3 | Study material | Structuring | 8 |
| 4 | Tasks in textbook | Structuring | 5 |
| 5 | Comics | Structuring | 9 |
| 6 | Stories, novels | Structuring | 6 |
| 7 | Poems | Structuring | 8 |
| 8 | Newspapers for children, textbooks and encyclopaedias | Evaluating | 6 |
| 9 | Advertising material | Evaluating | 5 |

From Janssen *et al.* (2000), with permission.

without taking into account the fact that each target is based on a different number of items and then correcting for this inequality. In our second analysis, we use a dichotomous version of the targets. In all of our analyses, both targets and respondents are considered to be taken from a random population, in agreement with Van den Noortgate *et al.* (2005), even though such a choice remains open to debate in the psychometric community.

## 3. Cross-classification multilevel models in psychometry

Suppose a set of items has been offered to a group of pupils and for each pupil the correctness of the responses has been recorded. These responses will usually vary, partly randomly, partly systematically, due to a difference in the ability of pupils to respond to an item correctly, as well as due to a difference in the difficulty of the items. Although one often encounters dichotomous items (e.g. correct/incorrect), item responses may also be categorical (e.g. yes/perhaps/no), and an appropriate aggregation can convert them into quasi-continuous responses. Let us refer to such aggregations as targets. We will focus first on such targets and hence on continuous outcomes. In the analysis of the case study, we will comment on issues stemming from targets that are made up of variable numbers of items. We will extend the methodology from the continuous case to the binary case in the second part of the analysis.

If both targets as well as people could be regarded as random samples from a population of targets and a population of people, one can define a random residual for both people and targets. Because people and targets are in a non-hierarchical relationship, classical methodology for hierarchical models is in need of extension. In the multilevel literature, such a model is therefore often referred to as a cross-classification model or crossed random-effects model (Goldstein, 1987; Raudenbush, 1993). Note that there is usually only one observation for each target–person combination.

## 4. Methodology for Gaussian outcomes

Let us consider a continuous random variable $Y_{ij}$ and a general model with two random factors $\alpha_i$ for participants ($i = 1, \ldots, I$) and $\beta_j$ ($j = 1, \ldots, J$) for targets, assumed to be crossed with each other. Further, we assume $\alpha_i$ and $\beta_j$ to follow $N(0, \sigma_\alpha^2)$ and $N(0, \sigma_\beta^2)$ distributions, respectively. The residual errors are assumed to follow a

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



$N(0, \sigma^2)$ distribution. All random terms are assumed to be independent of each other. The model can be written as

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}. \tag{1}$$

The parameters of interest are the variances $\sigma_\alpha^2$ and $\sigma_\beta^2$ of the random effects and the residual variance $\sigma^2$. We will now briefly describe the general conditional linear mixed model (Section 4.1) and then focus on the particular situation of crossed random effects (Section 4.2).

### *4.1. Conditional linear mixed models*

Conditional linear mixed models were used by Verbeke *et al.* (2001) to analyse hierarchical data without the need to specify time-independent effects such as the effect of baseline covariates that are assumed to have a constant effect over time.

Consider the general linear mixed-effects model, of which (1) is a special case

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{b} + \varepsilon, \tag{2}$$

where $\boldsymbol{\beta}$ corresponds to the fixed part of the model, $\mathbf{b}$ to the random part and the errors $\varepsilon$ are assumed to be normally distributed with zero mean and variance matrix equal to $\sigma^2 I$. Typically, the random effects $\mathbf{b}$ are assumed to be zero-mean normally distributed.

Verbeke *et al.* (2001) conceived conditional linear mixed-effects models to consist of two steps. In the first step, they conditioned on sufficient statistics for the cross-sectional component of the model. In order to proceed, they rewrote the general model

$$\mathbf{Y}_i = \mathbf{1}_{n_i} b_i^* + X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \varepsilon_i, \tag{3}$$

where the matrices $X_i$ and $Z_i$ are appropriate fixed-effects and random-effects design matrices and $\boldsymbol{\beta}$ and $\mathbf{b}_i$ are the corresponding parameter vectors.

The component $b_i^*$ groups all cross-sectional components, considered to be nuisance factors in this approach, and combining, for example, random intercepts and time-invariant effects of baseline covariates. Conditional linear mixed models now proceed in two steps. In a first step, we condition on sufficient statistics for the nuisance parameters $b_i^*$. In a second step, classical estimation procedures for nested random effects are used to estimate the remaining parameters in the conditional distribution of the $\mathbf{Y}_i$ given these sufficient statistics.

Conditional on the participant-specific parameters $b_i^*$ and $\mathbf{b}_i$ in (3), $\mathbf{Y}_i$ is normally distributed with mean vector $\mathbf{1}_{n_i} b_i^* + X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i$ and covariance matrix $\sigma^2 I_{n_i}$, from which it readily follows that $\bar{y}_i = \sum_j y_{ij}/n_i$ is sufficient for $b_i^*$. Further, the distribution of $\mathbf{Y}_i$, conditional on $\bar{y}_i$ and on the remaining participant-specific effects $\mathbf{b}_i$, is given by

$$\begin{aligned} f_i(\boldsymbol{y}_i|\bar{y}_i, \mathbf{b}_i) &= \frac{f_i(\boldsymbol{y}_i|b_i^*, \mathbf{b}_i)}{f_i(\bar{y}_i|b_i^*, \mathbf{b}_i)} \\ &= (2\pi\sigma^2)^{-(n_i-1)/2}\sqrt{n_i}\exp\left\{-\tfrac{1}{2\sigma^2}(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - Z_i\mathbf{b}_i)'\right. \\ &\quad \left.\times(I_{n_i} - \mathbf{1}_{n_i}(\mathbf{1}_{n_i}'\mathbf{1}_{n_i})^{-1}\mathbf{1}_{n_i}')(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - Z_i\mathbf{b}_i)\right\}. \end{aligned} \tag{4}$$

It now follows directly from matrix algebra (Seber, 1984) that (4) is proportional to

$$(2\pi\sigma^2)^{-(n_i-1)/2}\exp\left\{-\frac{1}{2\sigma^2}(A_i'\boldsymbol{y}_i - A_i'X_i\boldsymbol{\beta} - A_i'Z_i\mathbf{b}_i)'(A_i'A_i)^{-1}(A_i'\boldsymbol{y}_i - A_i'X_i\boldsymbol{\beta} - A_i'Z_i\mathbf{b}_i)\right\} \tag{5}$$

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



for any set of $n_i \times (n_i - 1)$ matrices $A_i$ of rank $n_i - 1$ which satisfy $A_i' \mathbf{1}_{n_i} = 0$. This shows that the conditional approach is equivalent to transforming each vector $\mathbf{Y}_i$ orthogonal to $\mathbf{1}_{n_i}$. If we now also require the $A_i$ to satisfy $A_i' A_i = I_{(n_i-1)}$, the transformed vectors $A_i' \mathbf{Y}_i$ satisfy

$$\begin{aligned} \mathbf{Y}_i^* \equiv A_i' \mathbf{Y}_i &= A_i' X_i \boldsymbol{\beta} + A_i' Z_i \mathbf{b}_i + A_i' \boldsymbol{\varepsilon}_i \\ &= X_i^* \boldsymbol{\beta} + Z_i^* \mathbf{b}_i + \boldsymbol{\varepsilon}_i^*, \end{aligned} \tag{6}$$

where $X_i^* = A_i' X_i$ and $Z_i^* = A_i' Z_i$ and where the $\boldsymbol{\varepsilon}_i^* = A_i' \boldsymbol{\varepsilon}_i$ are normally distributed with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I_{n_i - 1}$.

Model (6) is a linear mixed model, but with transformed data and covariates, such that the only parameters still in the model are the hierarchical effects and the residual variance. Hence, the second step in fitting conditional linear mixed models is to fit model (6) using maximum likelihood or restricted maximum likelihood methods. Note that once the transformed responses and covariates have been calculated, standard software for fitting linear mixed models (e.g. SAS procedure MIXED) can be used for the estimation of all parameters in model (6). This implies, as mentioned in the Introduction, that purely cross-sectional effects are removed from the set of estimable parameters.

Note that the proposed method is, in effect, very similar to restricted maximum likelihood (REML) estimation in the classical linear mixed model, where the variance components are estimated after transforming the data such that the fixed effects vanish from the model. As shown by Harville (1974, 1977) and by Patterson and Thompson (1971), and as discussed in Verbeke and Molenberghs (2000, Section 5.3.4), the REML estimates for the variance components do not depend on the selected transformation, and no information on the variance components is lost in the absence of information on the fixed effects. It has been shown by Verbeke *et al.* (2001) that similar properties hold for inferences obtained from conditional linear mixed models; that is, it was shown that results do not depend on the selected transformation $\mathbf{Y}_i - A_i' \mathbf{Y}_i$ and that no information is lost on the average, nor on the participant-specific effects, from conditioning on sufficient statistics for the cross-sectional components $b_i^*$ in the original model.

### *4.2. Models for crossed random effects*

Let us return to model (1) and apply conditional linear mixed model ideas to it. It is convenient to rewrite our model in vectorized form. To this end, group the outcomes into a vector $\mathbf{Y}$, with the second index varying more rapidly than the first one, and write (1) as

$$\mathbf{Y} = \mu \mathbf{1} + Z(\alpha) \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_I \end{pmatrix} + Z(\beta) \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_I \end{pmatrix} + \boldsymbol{\varepsilon}, \tag{7}$$

where the design matrices, using Kronecker products, are $Z(\alpha) = I_I \otimes \mathbf{1}_J$ and $Z(\beta) = \mathbf{1}_I \otimes I_J$, respectively. The dimensions of these are $IJ \times I$ and $IJ \times J$, respectively. We will now apply the conditional linear mixed model idea twice, once to remove the $\alpha_i$ and once to remove the $\beta_j$.

First, focusing on removal of the $\alpha_i$ effects, we need to construct a matrix $A$ with dimensions $IJ \times J(I - 1)$, such that $A' Z(\alpha) = \mathbf{0}$ and $A' A = I$. Assuming such a matrix has

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



been constructed, we obtain a transformed response vector

$$\mathbf{Y}^* = A'\mathbf{Y} = A'\mu\mathbf{1} + A'Z(\alpha)\begin{pmatrix}\alpha_1\\ \vdots \\ \alpha_I\end{pmatrix} + A'Z(\beta)\begin{pmatrix}\beta_1\\ \vdots \\ \beta_J\end{pmatrix} + A'\varepsilon = Z^*(\beta)\boldsymbol{\beta} + \varepsilon^*, \quad (8)$$

where $\varepsilon^*$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I_{J-1}$.

Let us now find a matrix $A$ such that $Z^*(\alpha) = A'Z(\alpha) = \mathbf{0}$ or, equivalently, $A'(I_I \otimes \mathbf{1}_J) = \mathbf{0}$. In this case and due to the specific format of $Z(\alpha)$, the matrix $A'$ can be written as $A_1' \otimes A_2'$. By using the fact that $(A_1' \otimes A_2')(I_I \otimes \mathbf{1}_J) = (A_1' I_I) \otimes (A_2' \mathbf{1}_J)$, we need to find a matrix $A_2'$ such that $A_2' A_2 = I_{J-1}$. To this end, let us define

$$A_{2,ij} = \begin{cases} 0, & 1 + j < i, \\ -j/\sqrt{j + j^2}, & 1 + j = i, \\ 1/\sqrt{j + j^2}, & 1 + j > i. \end{cases}$$

In particular, if we use $A_1' = I_I$, this implies

$$Z^*(\alpha) = A'Z(\alpha) = (A_1' \otimes A_2')(I_I \otimes \mathbf{1}_J) = (A_1' I_I) \otimes (A_2' \mathbf{1}_J) = I_I \otimes \mathbf{0}_{(J-1)\times 1} = \mathbf{0}_{I(J-1)\times I}$$

and the resulting model contains only one random factor $\beta$.

Second, with entirely similar logic, we need to apply a different transformation, $B$ say, to (7) to eliminate $\beta_j$ (details are omitted). We obtain a second conditional linear mixed model

$$\mathbf{Y}^{**} = Z^{**}(\alpha)\boldsymbol{\alpha} + \varepsilon^{**}. \quad (9)$$

Note that, just as $Z^*(\beta)$ has been removed by the transformation $A$, now $Z^{**}(\alpha)$ has been removed. Further, $\varepsilon^{**} \sim N(0, \sigma^2 I_{I-1})$, and hence the residual error variance component occurs in both (8) and (9), while the random effects occur in just one of these models.

## 5. Methodology for the binary case

We will start by introducing the model that will be used to analyse the binary version of the psychometric study, described in Section 2. Let us consider

$$\text{logit}[\Pr(Y_{ij} = 1 | a_i, b_j)] = \mu + a_i + b_j,$$

where $a_i$ and $b_j$ represent the random effects corresponding to person $i$ and item $j$, respectively. In addition, $Y_{ij}$ is a binary response with two possible outcomes: 1 for success and 0 for failure.

Of course, the effect of other covariates can be explored by adding extra terms to the model. Here, we will focus on models with two random effects.

It is useful to consider the specific case of matched pairs, that is the case of two items and two participants, and combine this setting with pseudolikelihood estimation.

This procedure has to be implemented in two stages. First, consider all possible pairs of items $(j, j')$. Therefore, there will be $\binom{J}{2}$ of such pairs within person $i$, where $J$ is the total number of items.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



If this strategy is repeated for all $I$ participants, we will then have a derived set of data containing $\binom{J}{2} \times I$ observations. Then, conditional logistic regression is applied by considering each pair as a unit. The resulting variance is $2\sigma_b^2$ and the variance of this last estimator can be obtained by means of the sandwich estimator $\text{var}(\sigma_b^2)(J-1)$. For more details about pseudolikelihood methods see Aerts, Geys, Molenberghs, and Ryan (2002).

Finally, in order to estimate the variance and its standard error corresponding to the factor item, a complete symmetric situation is considered where $\binom{I}{2}$ pairs of observations are constructed by using all possible participant pairs $(i, i')$. Then, $\text{var}(\sigma_a^2)(I-1)$ gives an estimator of the variance of the point estimator.

In the next sections, we will analyse the case study introduced in Section 2.

## 6. Analysis of a continuous outcome in the case study

To fit model (1) to the data described in Section 2, we will utilize conditionally derived models (8) and (9). First, we will ignore the fact that the targets are made up of different numbers of items and we will propose a strategy to account for it. The models were fitted to the data of our case study by means of the SAS procedure MIXED. Macros constructing the matrices $A$ and $B$, applying the orthogonal transformations before applying the SAS procedure MIXED, are available from the authors' website. In Table 2, the estimated values are displayed. The parameters, obtained with (9) and (8) are labelled 'Transf. 1' and 'Transf. 2', respectively.

**Table 2.** Parameter estimates (standard errors) for the conditional linear mixed-effects model, fitted to the psychometric data

| Effect | Parameter | Transf. 1 | Transf. 2 | Combined | Standard error |
|---|---|---|---|---|---|
| Person | $\hat{\sigma}_\alpha^2$ | 1.3634 | – | 1.3634 | 0.0017 |
| Target | $\hat{\sigma}_\beta^2$ | – | 2.2155 | 2.2155 | 0.0040 |
| Residual | $\hat{\sigma}^2$ | 0.6704 | 0.7477 | 0.7090 | 0.0030 |

Note that $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\beta^2$ are obtained from just one of the models, while the residual variance is estimated twice.

Since we are considering two conditional linear mixed models to estimate each of the two random-effects variances in turn, as a byproduct, we obtain two consistent estimates of the residual variance. In principle, we could simply select one of the two. However, in doing so, we would leave some information underused. Noting that each convex combination of the two is a sensible candidate for a combined estimate, a natural choice is to take the average. This also allows us to properly estimate the precision of the residual variance. Precisely, define $\hat{\sigma}^2 = (\hat{\sigma}_{(1)}^2 + \hat{\sigma}_{(2)}^2)/2$, producing an overall estimate of the residual variance. To obtain its standard error, let us proceed as follows. Construct

$$\hat{B} = \sum_{k=1}^{2} (\hat{\sigma}_{(k)}^2 - \hat{\sigma}^2)^2,$$

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



which reduces to $\hat{B} = (\hat{\sigma}^2_{(1)} - \hat{\sigma}^2_{(2)})^2/2$ and $\hat{W} = (v_{(1),11} + v_{(2),11})/2$, with $v_{(k),11}$ the element of the covariance matrix of the covariance parameters for model $k = 1, 2$. The covariance matrix of the covariance parameters can then be written as $\hat{V} = \hat{W} + \hat{B}$. Table 2 shows the final numerical results of the estimation process in the last two columns.

### *6.1. Discussion and scope of results*

Table 2 shows that the expected score of a pupil on a target varies over people and especially over targets. Both variance components are highly significant ($p < .0001$) using a Wald test, whether or not one corrects for the fact that the null hypothesis lies on the boundary of the parameter space (Verbeke & Molenberghs, 2000, Section 6.3). Although pupils and targets explain a very large part of the total variability, 32 and 52%, respectively, the residual variability is substantial and statistically highly significant as well ($p < .0001$). This indicates that the score for a pupil on a target may deviate from the score that is expected based on the person ability and the target difficulty. It implies that there is some interaction between people and targets.

Our model contains random effects only; no fixed effects are present apart from an intercept that is removed in the conditioning process. This means that both person abilities and target difficulties are assumed to be independently and identically drawn from a normal distribution with variances as in Table 2. Further exploration of these person abilities and target difficulties can be done by studying the empirical Bayes estimates of the $\alpha_i$ and $\beta_j$. In many practical situations, however, one may want to explain, for example, differences in target difficulties based on a priori grounds, say, using target-level covariates. The same might be true at the level of the pupil. Such person-level covariates may be continuous (e.g. age) or categorical (e.g. sex). Similarly, target characteristics (e.g. number of subtasks, as discussed below, or the type of problem) can often be assumed to influence the target difficulty. It is also possible that person-by-target characteristics have an effect.

Of course, model (7) can easily be extended by including such person, target or interaction effects as covariates. This is conveniently done by replacing $\mu\mathbf{1}$ in (7) by a full fixed-effects design, $X\boldsymbol{\beta}$. The inclusion of covariates can be based on prior beliefs of the researcher about effects of these characteristics, but may also be a tool to explore possible relations. The use of the cross-classification model complemented with target and person predictors yields a flexible predictive and explanatory approach, as it includes error terms on both sides. For instance, the targets in our example are characterized by a combination of text type and the level of processing. One could include one or both variables in the model to explore whether they explain part of the variance between targets. Since this would be a diversion from our methodological development, we have chosen not to discuss this further.

A person covariate of particular interest is a person group. For instance, pupils can be grouped into schools. When the groups are seen as randomly drawn from a population of groups, the group effects can be modelled as random rather than as fixed effects. For the cross-classification model discussed above, this would result in a model with crossed as well as nested random effects. Fox and Glas (2001, 2003) applied these types of model in a similar context and they used Markov Chain Monte Carlo (MCMC) methods for parameter estimation. Although the conditional linear mixed model approach could still be used, the discussion of this extension of the simple cross-classification model is beyond the scope of this article.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



### *6.2. Unequal number of items per target*

From Table 1 we see that the number of items per target is variable. This is bound to occur whenever a combined score, such as the targets used here, is the subject of analysis. In some cases, unlike here, the number of items per targets varies widely and if one insists on retaining stability of variances at the item level, rather than at the target level, then one can proceed as follows. Extend (1) by denoting $Y_{ijk}$ as the response by person $i = 1, \ldots, I$ on item $k = 1, \ldots, K_j$ contributing to target $j = 1, \ldots, J$ and write the item level model

$$Y_{ijk} = \tilde{\mu} + \alpha_i + \beta_j + \tilde{\varepsilon}_{ijk}, \tag{10}$$

with distributions $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_j \sim N(0, \sigma_\beta^2)$, and $\tilde{\varepsilon}_{ijk} \sim N(0, \sigma^2)$. Then, the derived target level model is

$$Y_{ij} = \frac{\sum_{k=1}^{K_j} Y_{ijk}}{K_j} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \tag{11}$$

with unaltered distributions for the random effects and

$$\varepsilon_{ij} \sim N(0, \sigma^2/K_j). \tag{12}$$

In other words, this model is similar to (1), except for a heteroscedastic measurement error variance. Next, we need to apply transformations *A* and *B,* as in Section 4.2, but with a slightly different requirement. For example, we would still have $\mathbf{Y}_i \rightarrow \tilde{A}_i'\mathbf{Y}_i$, requiring $\tilde{A}_i$ to be orthogonal to the unit vector, but with condition $\tilde{A}_i'\Lambda_i\tilde{A}_i = I_{(n_i-1)}$, where $\Lambda_i$ is a diagonal matrix with *j*th diagonal element equal to $1/K_j$. Then, after transformation, the residual errors are once more zero-mean normally distributed with covariance $\sigma^2 I_{(n_i-1)}$ and the only programming required is the implementation of this alternative transformation. For the *B* transformation, where the data are grouped by target *j* rather than by persons *i*, the only modification required is to multiply all elements of *B* as described in Section 4.2 by $\sqrt{K_j}$.

Doing so produces $\hat{\sigma}_\beta^2 = 29.11$ ($SE = 0.42$), $\hat{\sigma}_\alpha^2 = 13.08$ ($SE = 0.08$) and $\hat{\sigma}^2 = 0.9506$ ($SE = 0.0006$). The numerical difference in random-effects variances is, of course, due to the fact that the variability is now at item level rather than at target level, and hence the inclusion of the factor $K_j$ in (12) changes the balance between the random-effects variances and the measurement error variance. A further reason why the results differ here, in comparison to those obtained at the target level is that we now properly account for the different number of items per target, whereas before this aspect was ignored.

### *6.3. Simulation study*

We conducted a simulation study to evaluate the performance of the conditional linear mixed model for crossed random effects. The design of the simulation study was carried out under different settings to investigate the impact of the number of participants and number of items on the performance.

First, we generated data where the true parameters were set equal to the estimates obtained from the analysis done previously. Five hundred simulation data sets were generated. Other, additional, settings were also used to study changes on the variances of the random effects. They were defined in the following way:

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



Setting 1. $\sigma_\alpha^2 = 1.3634; \sigma_\beta^2 = 2.2155$; $\sigma^2 = 0.7090$; $I = 5$, 10, 20, 30; $J = 10$, 20, 50, 100.

Setting 2. $\sigma_\alpha^2 = 0.50$ to 8.50 by 0.5; $\sigma_\beta^2 = 2.2155$; $\sigma^2 = 0.7090$; $J = 20$, 50, 100; $I = 10$.

Setting 3. $\sigma_\alpha^2 = 1.3634; \sigma_\beta^2 = 0.50$ to 8.50 by 0.5; $\sigma^2 = 0.7090$; $J = 20$, 50, 100; $I = 10$.

We report the results of setting 1 in Table 3; bias and relative bias were calculated by taking the average of $\hat{\sigma}^2 - \sigma^2$ from 500 replicates. Mean (*SE*) denotes the average of the estimated standard errors of the estimates. The 95% confidence interval coverage probabilities are included as well.

From this table it can be seen that the method performs well in most cases. The relative bias regarding the estimation of $\sigma_\alpha^2$ decreases when the number of participants increases. However, it is never larger than 0.06. An identical situation can be observed for $\sigma_\beta^2$, where the relative bias decreases when the number of items increases, being always smaller than 0.04. This similarity is, of course, to be expected.

The 95% coverage probabilities are all rather high and even in the most unfavourable situation, that is a low number of participants as well as of items, the coverage probabilities are 98.6 and 90.2% for the variance of the participants and items, respectively. It is important to point out that our case study contains nine items and 500 participants making the simulation with 10 items and 100 participants the closest to it. In this setting we obtained a coverage of 100 and 96.4% for each of the crossed random effects.

To explore the impact of the magnitude of the variances on the estimation, we performed simulations by fixing one of the variances and varying the other, between 0.5 and 8.5. The variances were fixed to be 1.3634 and 2.2155, according to the values obtained from the data set. We conducted the simulation study for a fixed number of 10 items and for four different choices for the number of participants, i.e. 10, 20, 50 and 100.

The numerical results are graphically displayed to facilitate interpretation. Figure 1 contains the results of the simulations with fixed variance of $\beta$ and Figure 2 displays equivalent results with fixed variance of $\alpha$. Each panel corresponds to a different number of participants. True values are plotted against estimated values, together with their confidence intervals. The dotted lines indicate the estimated values of $\alpha$.

These figures show that almost all points virtually fall on the dotted line, indicating a high agreement between true and estimated values. As can be expected, the size of the confidence intervals increases with variance and decreases with number of participants.

## 7. Analysis of binary outcome in the case study

To fit the model described in Section 5 to the binary outcomes from the study introduced in Section 2, we start by restructuring the data in an appropriate way. To this end, we construct two new data sets. The first one contains $\binom{I}{2} \times J$ observations where the rows correspond to all possible pairs that can be formed out of the items, and $I$ is the total number of persons in the original data set. Similarly, the second data set contains $\binom{I}{2} \times I$ observations where pairs are now formed from participants. This is common practice in a so-called pairwise pseudolikelihood context (Aerts *et al.*, 2002).

We developed an SAS macro to perform these operations in an automated fashion. After having arranged the data in a more convenient way, we fit two conditional logistic models by means of the SAS procedure NLMIXED. The code is available from the authors' website.

**Table 3.** Results of the simulation study

| | Participants | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Items | Parameters | $\sigma_\alpha^2$ | $\sigma_\beta^2$ | $\sigma^2$ | $\sigma_\alpha^2$ | $\sigma_\beta^2$ | $\sigma^2$ | $\sigma_\alpha^2$ | $\sigma_\beta^2$ | $\sigma^2$ |
| 5 | Estimate | 1.281 | 2.125 | 0.769 | 1.287 | 2.126 | 0.771 | 1.289 | 2.132 | 0.772 |
| | Bias | 0.081 | 0.089 | 0.060 | 0.076 | 0.088 | 0.062 | 0.073 | 0.083 | 0.064 |
| | Rel. bias | 0.060 | 0.040 | 0.084 | 0.055 | 0.040 | 0.088 | 0.054 | 0.037 | 0.089 |
| | Mean (*SE*) | 0.847 | 0.339 | 0.018 | 0.536 | 0.213 | 0.011 | 0.381 | 0.150 | 0.008 |
| | 95% coverage | 98.6 | 90.2 | | 99.2 | 94.0 | | 99.8 | 95.8 | |
| 10 | Estimate | 1.290 | 2.133 | 0.772 | 1.287 | 2.135 | 0.772 | 1.286 | 2.134 | 0.772 |
| | Bias | 0.072 | 0.082 | 0.063 | 0.075 | 0.079 | 0.063 | 0.076 | 0.081 | 0.063 |
| | Rel. bias | 0.053 | 0.037 | 0.088 | 0.055 | 0.035 | 0.089 | 0.056 | 0.036 | 0.090 |
| | Mean (*SE*) | 0.568 | 0.242 | 0.012 | 0.360 | 0.150 | 0.008 | 0.254 | 0.105 | 0.006 |
| | 95% coverage | 99.4 | 91.2 | | 99.8 | 94.2 | | 100 | 96.4 | |
| 20 | Estimate | 1.289 | 2.131 | 0.772 | 1.287 | 2.132 | 0.772 | 1.289 | 2.134 | 0.773 |
| | Bias | 0.074 | 0.084 | 0.063 | 0.076 | 0.082 | 0.064 | 0.077 | 0.081 | 0.064 |
| | Rel. bias | 0.054 | 0.038 | 0.089 | 0.055 | 0.037 | 0.089 | 0.056 | 0.036 | 0.090 |
| | Mean (*SE*) | 0.390 | 0.171 | 0.008 | 0.247 | 0.106 | 0.006 | 0.175 | 0.074 | 0.004 |
| | 95% coverage | 99.8 | 95.4 | | 100 | 96.8 | | 100 | 97.4 | |
| 30 | Estimate | 1.288 | 2.133 | 0.772 | 1.287 | 2.134 | 0.773 | 1.286 | 2.135 | 0.773 |
| | Bias | 0.074 | 0.081 | 0.064 | 0.075 | 0.080 | 0.064 | 0.076 | 0.079 | 0.064 |
| | Rel. bias | 0.054 | 0.036 | 0.089 | 0.055 | 0.036 | 0.090 | 0.056 | 0.035 | 0.090 |
| | Mean (*SE*) | 0.316 | 0.140 | 0.007 | 0.200 | 0.087 | 0.004 | 0.141 | 0.061 | 0.003 |
| | 95% coverage | 99.8 | 97.0 | | 100 | 96.6 | | 100 | 96.2 | |



Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



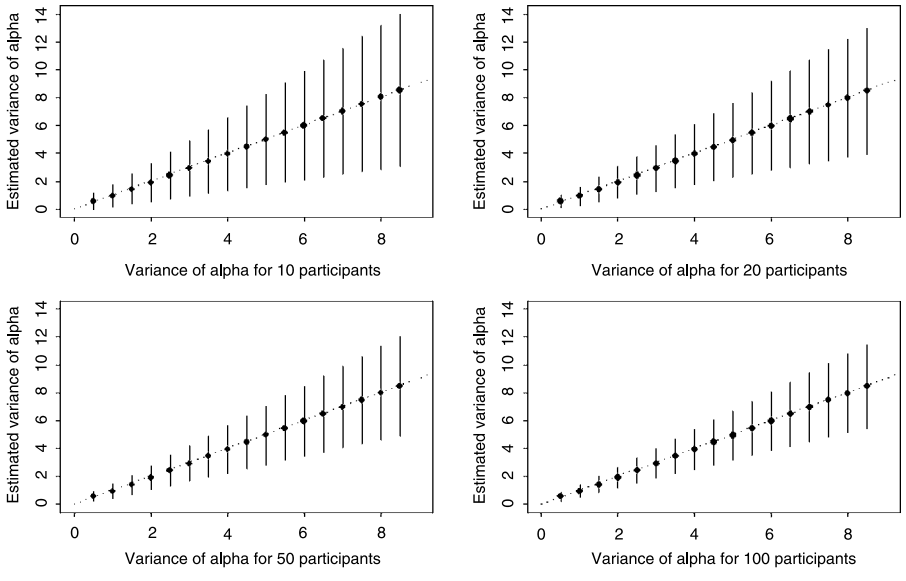



**Figure 1.** Simulation results for variance of beta equal to 2.2155 and variance of alpha varying between 0.5 and 8.5. Each panel corresponds to different numbers of participants. The vertical lines indicate the size of the 95% confidence intervals.

In Table 4, the estimated values of $\sigma_a^2$ and $\sigma_b^2$ and their standard errors are presented using the fact that procedure NLMIXED produces an estimate of $2\sigma_a^2$ and $2\sigma_b^2$. By mere recoding, direct estimates of $\sigma_a^2$ and $\sigma_b^2$ can be obtained.

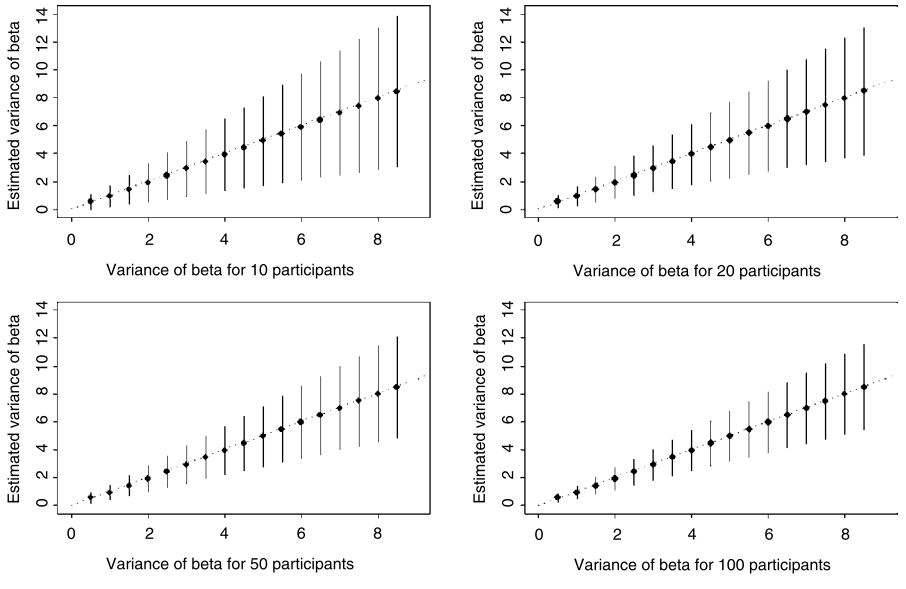



**Figure 2.** Simulation results for variance of alpha equal to 1.3634 and variance of beta varying between 0.5 and 8.5. Each panel corresponds to different numbers of participants. The vertical lines indicate the size of the 95% confidence intervals.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



**Table 4.** Psychometric study: parameter estimates (standard errors) for the conditional logistic mixed-effects model, fitted to the psychometric data

| Effect | Parameter | Estimates | Standard error |
|---|---|---|---|
| Person | $\hat{\sigma}_a^2$ | 0.0795 | 0.927 |
| Item group | $\hat{\sigma}_b^2$ | 0.0405 | 0.212 |

Note that these variances have the same meaning as their counterparts in the continuous case, in spite of the fact that we now use binary outcomes. This is because the variances characterize the continuous random-effect distributions at the item and participant levels, respectively. Of course, the numerical values cannot be compared across analyses because in the continuous case we consider targets whereas here the focus is on items.

An important issue when using the NLMIXED procedure is convergence. First, different initial values of the parameters can be tried if convergence fails with the default option. Second, different optimization techniques can be chosen and some of them are faster to reach convergence. In our case, we have tried different choices of initial values of the parameters and different optimization techniques in order to obtain convergence. However, once convergence is reached it seems that the method is quite stable with respect to perturbation of initial values.

From the numerical values obtained after fitting the models, it appears that the variability between participants is larger than the variability between items, but this conclusion should be used with caution given that the Wald test fails to reject the null hypotheses due to the large standard errors.

It is clear that since model parameters between the continuous and binary cases have a different meaning, estimates between this and the previous section are not directly comparable.

## 8. Concluding remarks

We have proposed an estimation method for the variance components of a crossed random-effects model when only one observation is available in each cross-classified level. We have illustrated the methodology using data from a psychometric study. Of course, there is no restriction on our methodology being applied in different scientific fields. To estimate the variance of the error and the standard deviation of these variances, we proposed an approach of combining the values obtained from the fit of two simple models after an appropriate transformation of the response and the effects. We implemented our approach by means of an SAS macro.

We do not claim our method to be uniformly better than conventional maximum likelihood methods. While our method can lead to computational economy, modern computational tools do allow the researcher to use standard MLE in rather broad settings (Raudenbush, 1993). This is particularly true in the continuous case, but less so in the binary case, where MLE for random-effects models can be computationally quite cumbersome (Molenberghs & Verbeke, 2005).

In addition, in the conditional approach, the cross-sectional structure does not require (correct) specification, which leads to bias reduction. Of course, this comes at the price of the inability to make inferences about the cross-sectional structure. In this

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



sense, the work of Miyazaki and Raudenbush (2000) exhibits an important instance where it would not make sense to consider our approach. Further, as is the case in many other areas in statistics, bias reduction may come with some efficiency loss. This result is supported by our simulations.

The estimation of the variance of the random effects in the binary case has been carried out by using conditional logistic regression together with pseudolikelihood ideas. We have established the relationship between our approach and the framework of matched pairs.

We have conducted several simulation studies of the proposed method under a variety of circumstances, e.g. different numbers of participants, items and a range of values for the variance of $\alpha$ and $\beta$ for the continuous case. Our method performs quite well for the estimation of the variance components in most of the cases. Some extensions of this method can easily be carried out; for example, fixed covariates can be included in the model without major changes to the general structure of the programs.

We have proposed an alternative estimation method for the variance component of a crossed random-effect model when the response is binary and, in addition to that, only one observation is available in each crossed classified level.

## Acknowledgements

We wish to thank Bijzonder Onderzoeksfonds of the Limburgs Universitair Centrum. We acknowledge support from Interuniversity Attraction Poles Program P5/24-Belgian State-Federal Office for Scientific, Technical and Cultural Affairs.

## References

Aerts, M., Geys, H., Molenberghs, G., & Ryan, L. (2002). *Topics in modelling of clustered data*. London: Chapman and Hall.

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

De Boeck, P., Daems, F., Meulders, M., & Rymenams, R. (1997). *Onwikkeling van een toets voor de eindtermen begrijpend lezen* [Construction of a test for the educational target of reading comprehension]. Leuven/Antwerpen (Belgium): University of Leuven/University of Antwerpen.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, *6*, 341–353.

Fahrmeir, L., & Tutz, C. (2001). *Multivariate statistical modeling based on generalized models* (2nd ed.). New York: Springer.

Fox, J., & Glas, C. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271–288.

Fox, J., & Glas, C. (2003). Bayesian modelling of measurement error in predictor variables using item response theory. *Psychometrika*, *68*, 169–191.

Goldstein, H. (1987). Multilevel covariance components models. *Biometrika*, *74*, 430–431.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, *61*, 383–385.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*, 320–340.

Janssen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics*, *25*, 285–306.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society



Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods*, *5*, 44–63.

Molenberghs, G., & Verbeke, C. (2005). *Discrete longitudinal data*. New York: Springer.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*, 545–554.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*, *18*, 321–349.

Seber, G. A. F. (1984). *Multivariate observations*. New York: Wiley.

Van den Noortgate, W., De Boeck, P., & Meulders, M. (2005). Cross-classification multilevel logistic models in psychometry. *Journal of Educational and Behavioral Statistics*, *28*, 369–386.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

Verbeke, G., Spiessens, B., & Lesaffre, E. (2001). Conditional linear mixed models. *American Statistician*, *55*, 25–34.

*Received 29 November 2004; revised version received 9 March 2006*