

A Nonlinear Mixed Model Framework for Item Response Theory

Frank Rijmen, Francis Tuerlinckx, Paul De Boeck, and Peter Kuppens
Katholieke Universiteit Leuven

Mixed models take the dependency between observations based on the same cluster into account by introducing 1 or more random effects. Common *item response theory* (IRT) models introduce latent person variables to model the dependence between responses of the same participant. Assuming a distribution for the latent variables, these IRT models are formally equivalent with *nonlinear mixed models*. It is shown how a variety of IRT models can be formulated as particular instances of nonlinear mixed models. The unifying framework offers the advantage that relations between different IRT models become explicit and that it is rather straightforward to see how existing IRT models can be adapted and extended. The approach is illustrated with a self-report study on anger.

Mixed models (see, e.g., Diggle, Heagerty, Liang, & Zeger, 2002; Goldstein, 1995; Longford, 1993; Verbeke & Molenberghs, 2000) are a collection of statistical tools that are well suited for analyzing clustered data, such as, for example, data from students nested within schools or repeated measurement data (measurements nested within participants). Modeling clustered data under a model that assumes independent observations is inappropriate because the observations on the subunits (students, measurements) of the same unit (school, participant) tend to be more homogeneous than the observations on subunits of different units. The heterogeneity between units can be taken into account by assuming that (some of) the parameters of the model follow some random distribution over the population of units. Hence, (some of) the parameters of the model are random variables, and the model is a *random effects model*, or *mixed model*.

Frank Rijmen, Francis Tuerlinckx, Paul De Boeck, and Peter Kuppens, Department of Psychology, Katholieke Universiteit (K.U.) Leuven, Leuven, Belgium.

An Appendix is available in the online version of this article in the PsycARTICLES database.

Frank Rijmen was supported by the Fund for Scientific Research Flanders. Francis Tuerlinckx was supported by National Science Foundation Grant SES00-84368 and by the Fund for Scientific Research Flanders. Peter Kuppens was supported by Grant GOA/2000/02 from K.U. Leuven.

Correspondence concerning this article should be addressed to Frank Rijmen, Department of Psychology, K.U. Leuven, Tiensestraat 102, B-3000 Leuven, Belgium. E-mail: frank.rijmen@psy.kuleuven.ac.be

The *random effects* represent the unit effects. The parameters that are not assumed to be random realizations from a distribution, and hence are not considered to be unit specific, are called *fixed effects*.

Mixed models and related methods were first developed in the context of analysis of variance and regression analysis, leading to the *linear mixed model*. Other commonly used terms are *multilevel models* (Goldstein, 1995), *hierarchical models* (Raudenbush & Bryk, 2002), and *random coefficient models* (Longford, 1993). More recently, *nonlinear mixed models* were also developed. A *nonlinear mixed model* is a model with random coefficients in which the fixed and/or random effects enter the model nonlinearly (Davidian & Giltinan, 1995). As we explain below, a subset of nonlinear mixed models is the class of *generalized linear mixed models* (McCulloch & Searle, 2001).

The main purpose of this article is to explain how *item response theory* (IRT) models can be conceptualized as nonlinear mixed models and to provide a framework for this conceptualization. There are four important assets of this approach. First, this conceptualization relates IRT to the broad statistical literature on mixed models. Second, applying the same framework to different IRT models can help in the understanding of their differences and commonalities. Third, using this framework, one can readily adapt or extend standard IRT models, so that a researcher can build his or her own model, customized to a specific scientific question or data set. Fourth, existing and newly formulated models can be estimated using soft-

ware for generalized linear and nonlinear mixed models.

The article is organized as follows. First, we present the mixed logistic regression model (Hedeker & Gibbons, 1994; Zeger & Karim, 1991) and show how different IRT models fit within this framework. It is explained how those IRT models can be categorized on the basis of which kind (or combination) of covariates the model takes into account: *item covariates*, *person covariates*, and *person-by-item covariates*. The mixed logistic regression model is presented for binary data. Then, models for polytomous data are considered. With respect to the latter, an additional distinction between models refers to how logits for the response variable are formed: *baseline-category logits*, *adjacent-category logits*, *continuation-ratio logits*, and *cumulative logits*.

Subsequently, the mixed logistic regression model is extended to incorporate IRT models that are nonlinear mixed models but do not belong to the family of generalized linear mixed models, unlike the mixed logistic regression model. These models have in common that multiplicative functions of parameters appear in the model equation (e.g., a product of a discrimination parameter and a latent trait). Then we discuss briefly some remaining IRT models that cannot be captured by the mixed logistic regression model or by our extension of it.

Next, statistical inference and software packages are discussed. The general models we present are illustrated with specific instantiations. In addition, a self-report study on anger is analyzed. The SAS code used to estimate these models with the SAS procedure NLMIXED is provided in Appendixes A and B, which are available in the online version of this article in the PsycARTICLES database.

Mixed Logistic Regression Model for Binary Data

In the mixed logistic regression model for binary data, the observations are assumed to be independent Bernoulli observations conditional on the covariates, the fixed effects, and the random effects. This conditional independence assumption is often referred to in the IRT literature as the assumption of *local stochastic independence*. The probability of success for each observation is modeled as

$$p(Y_{ni} = 1 | \mathbf{x}_{ni}, \mathbf{z}_{ni}, \boldsymbol{\beta}, \boldsymbol{\theta}_n) = \frac{\exp(\mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{ni}\boldsymbol{\theta}_n)}{1 + \exp(\mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{ni}\boldsymbol{\theta}_n)} \quad (1)$$

where Y_{ni} is the binary response variable for subunit i of unit n , $i = 1, \dots, I_n$; $n = 1, \dots, N$; \mathbf{x}_{ni} is a known P -dimensional covariate vector for the P fixed effects; \mathbf{z}_{ni} is a known Q -dimensional covariate vector for the Q random effects; $\boldsymbol{\beta}$ is the P -dimensional parameter vector of fixed effects; and $\boldsymbol{\theta}_n$ is the Q -dimensional parameter vector of random effects for unit n .

The distribution of the random effects $\boldsymbol{\theta}_n$ is commonly assumed to be multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N$ are assumed to be independent. The random effects have a mean of zero because they represent the deviations from the mean effects of the covariates, which are treated as fixed effects.

The marginal density of a response vector \mathbf{y}_n of length I_n is

$$p(\mathbf{y}_n | \mathbf{X}_n, \mathbf{Z}_n, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int \left\{ \prod_{i=1}^{I_n} \frac{\exp[y_{ni}(\mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{ni}\boldsymbol{\theta}_n)]}{1 + \exp(\mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{ni}\boldsymbol{\theta}_n)} \right\} N(\boldsymbol{\theta}_n | \boldsymbol{\Sigma}) d\boldsymbol{\theta}_n, \quad (2)$$

where \mathbf{X}_n and \mathbf{Z}_n are the $I_n \times P$ and $I_n \times Q$ design matrices for unit n of the fixed and random effects, respectively, with \mathbf{x}'_{ni} and \mathbf{z}'_{ni} as i th rows.

Furthermore, we define \mathbf{X} and \mathbf{Z} to be the supermatrices obtained from stacking the N \mathbf{X}_n and \mathbf{Z}_n matrices one below the other. This formulation will turn out to be useful to distinguish between several kinds of covariates.

The mixed logistic regression model belongs to the class of generalized linear mixed models (Clayton, 1996; Hedeker & Gibbons, 1994; McCulloch & Searle, 2001; Zeger & Karim, 1991). To clarify the latter, we briefly introduce the generalized linear mixed model in the following.

In a generalized linear model (McCullagh & Nelder, 1989) the observations are assumed to be independent realizations from an exponential family distribution, and the mean responses $\mu_i = E(Y_i | \mathbf{x}_i, \boldsymbol{\beta})$ are related to a linear predictor η_i via a link function $g(\cdot)$,

$$g(\mu_i) = \eta_i = \mathbf{x}'_i \boldsymbol{\beta}. \quad (3)$$

If the distribution is the normal distribution, and the link function $g(\cdot)$ is the identity link function, $g(\mu_i) = \mu_i$, the classical linear model is obtained.

Generalized linear models are a subset of nonlinear models:

1. Generalized linear models are nonlinear models, because the conditional mean is not a linear

combination of parameters, except when $g(\cdot)$ is the identity link, and the data are not necessarily normally distributed as in linear models.

2. Generalized linear models are a subset of nonlinear models, because the observations are distributed according to some exponential family distribution, and some function of the conditional mean is a linear combination of parameters.

In a generalized linear mixed model, the observations are assumed to be independent realizations from an exponential family distribution conditional on the random effects (in addition to the covariates and fixed effects). The mean responses $\mu_{ni} = E(Y_{ni} | \mathbf{x}_{ni}, \mathbf{z}_{ni}, \boldsymbol{\beta}, \boldsymbol{\theta}_n)$ are now related via a link function $g(\cdot)$ to a linear predictor that consists of both a fixed and a random part:

$$g(\mu_{ni}) = \eta_{ni} = \mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{ni}\boldsymbol{\theta}_n. \quad (4)$$

If the distribution is the normal distribution and $g(\cdot)$ is the identity link function, a linear mixed model is obtained.

The mixed logistic regression model is a generalized linear mixed model in which the observations are realizations from a Bernoulli distribution, $\mu_{ni} = \pi_{ni} = p(Y_{ni} = 1 | \mathbf{x}_{ni}, \mathbf{z}_{ni}, \boldsymbol{\beta}, \boldsymbol{\theta}_n)$, and the link function is the *logit function*,

$$L(\pi_{ni}) = \log \frac{\pi_{ni}}{1 - \pi_{ni}}.$$

Hence, an equivalent formulation for the mixed logistic regression model is

$$L(\pi_{ni}) = \eta_{ni} = \mathbf{x}'_{ni}\boldsymbol{\beta} + \mathbf{z}'_{ni}\boldsymbol{\theta}_n. \quad (5)$$

Two other commonly used link functions for binary data are the *probit link*, $\eta_{ni} = \Phi^{-1}(\pi_{ni})$, where $\Phi^{-1}(\cdot)$ is the cumulative standard normal distribution, and the *complementary log-log link*, $\eta_{ni} = \log[-\log(1 - \pi_{ni})]$ (McCullagh & Nelder, 1989).

In psychometrics, IRT models have been developed to model clustered binary data (polytomous data are discussed later). Commonly, in this context, the sub-units are the responses on the items, and the units are the participants (cf. repeated measurement data). In IRT models, one or more latent person variables are incorporated in the model to account for participant effects. If it is assumed that the latent variable(s) follow a distribution, the IRT model is formally equivalent to a nonlinear mixed model, where the latent

variables are the random effects (Agresti, Booth, Hobert, & Caffo, 2000; Kamata, 2001; Legler & Ryan, 1997).

For categorization purposes, we now introduce the following definitions:

1. *Item covariates*: A covariate is an item covariate if and only if the elements of the corresponding column of \mathbf{X} (and/or \mathbf{Z}) vary across items but are constant across persons.
2. *Person covariates*: A covariate is a person covariate if and only if the elements of the corresponding column of \mathbf{X} (and/or \mathbf{Z}) vary across persons but are constant across items.
3. *Person-by-item covariates*: A covariate is a person-by-item covariate if and only if the elements of the corresponding column of \mathbf{X} (and/or \mathbf{Z}) vary across both persons and items.

IRT models that are mixed logistic regression models for binary data can be classified on the basis of the kind of covariates taken into account in the model: item covariates, person covariates, and person-by-item covariates:

1. Examples of IRT models for binary data with only item covariates are the Rasch model (Rasch, 1960), the linear logistic test model (Fischer, 1973), and the random weights linear logistic test model (Rijmen & De Boeck, 2002).
2. Examples of IRT models in which the latent variable is regressed on person covariates include Adams, Wilson, and Wu (1997); Mislevy (1987); and Zwinderman (1991). These models incorporate item covariates as well as person covariates. Note that the weights of the latent regression on the latent variable are fixed effects and that the random effect corresponds with the error term of the latent regression.
3. The dynamic Rasch model (Verguts & De Boeck, 2000; Verhelst & Glas, 1993) and common models for differential item functioning are IRT models that incorporate both item covariates and person-by-item covariates.

Within the context of IRT, the generalized Rasch model of Adams and Wilson (Adams, Wilson, & Wang, 1997; Adams, Wilson, & Wu, 1997; Wu, Adams, & Wilson, 1998) closely resembles the mixed

logistic regression model of Equation 2, taking into account item and person covariates and, insofar as differential item functioning is concerned, also person-by-item covariates. The generalized Rasch model was developed outside the context of generalized linear mixed models, however. We now illustrate how different IRT models can be formulated as specific instances of the mixed logistic regression model of Equation 2.

Only Item Covariates

The Rasch model. Under the Rasch model (Rasch, 1960), the responses are assumed to be conditionally independent Bernoulli observations, with the probabilities of success modeled as follows:

$$p(Y_{ni} = 1 | \beta_i, \theta_n) = \frac{\exp(\theta_n + \beta_i)}{1 + \exp(\theta_n + \beta_i)}, \quad (6)$$

where β_i is the item parameter of item i , $i = 1, \dots, I$, and θ_n is the person parameter (ability) of person n , $n = 1, \dots, N$.

The person parameter θ_n is a latent variable. If the θ_n values are considered to be a random sample from a (commonly normal) distribution, then Equation 6 is a mixed logistic regression model with linear predictor

$$\eta_{ni} = \theta_n + \beta_i \quad (7)$$

(see Equations 2 and 5), where \mathbf{X}_n is an $I \times I$ identity matrix and \mathbf{Z}_n a vector of ones of length I for all n , so that \mathbf{X} consists of N stacked $I \times I$ identity matrices, and \mathbf{Z} is a column vector of ones of length $N \times I$, formed by stacking the N column vectors of ones of length I one below the other. Hence, in the Rasch model only the intercept θ_n is random. Furthermore, the model incorporates only item covariates, with fixed effects $\boldsymbol{\beta} = (\beta_1, \dots, \beta_I)'$.

The mean of the latent variable is constrained to zero, conforming to the convention to define the random effects as the deviations from the mean effect. In this particular case, the fixed intercept parameter is constrained to be zero as well, to ensure that the model is identifiable (otherwise, \mathbf{X} would not be of full column rank). The Rasch model is mainly a measurement model because it does not provide a substantive explanation for the data.

The linear logistic test model (LLTM). The LLTM (Fischer, 1973) is a Rasch model with a linear structure on the item parameters. Its specification as a mixed logistic regression model is completely analogous to the specification of a Rasch model as a mixed logistic regression model, except that the N identical

\mathbf{X}_n matrices no longer have to be identity matrices. Their columns represent item covariates (note that Fischer, 1973, used the term *weights* for the values of the covariates and the term *basic parameters* for the logistic regression weights). For example, an item covariate can represent the number of times a particular operation has to be performed to answer an item correctly. Furthermore, \mathbf{X} now does in general also contain a column of ones coding for the fixed intercept. Of course \mathbf{X} should be of full column rank to ensure the identifiability of the model.

The LLTM explains the responses by relating them to external item characteristics as represented in \mathbf{X}_n . This allows for explaining why some items are more difficult than others or why “yes” responses are more common for some items than for others.

The random weights linear logistic test model (RWLLTM). The RWLLTM (Rijmen & De Boeck, 2002) extends the LLTM in allowing that, in addition to the intercept, also one or more of the item covariates have a random effect. For example, a random effect for an item covariate that represents a cognitive operation to be carried out in some but not all of the items means that the cognitive operation involved in these items has a person-dependent effect on the probability of success, posing more difficulty for some persons than for others. The random effects represent the person-specific deviations from the mean (fixed) effect of the item characteristic. Because it introduces additional random effects, the RWLLTM is a multi-dimensional (each random effect corresponds to a dimension) extension of the unidimensional LLTM.

In the LLTM, the main effects of participants are taken into account by the random intercept, and the main effects of the item covariates by their fixed effects. However, interactions between item covariates and participants are not captured by the LLTM. In contrast, in the RWLLTM, the item covariates may have a person-dependent (random) effect, so that interactions between item covariates and participants can indeed be taken into account.

Casting the RWLLTM in terms of the mixed logistic regression model, we see that both \mathbf{X} and \mathbf{Z} consist of N stacked identical matrices (containing only item covariates). In contrast to the \mathbf{Z} matrix that specifies the Rasch model and the LLTM, \mathbf{Z} now contains a column for each item covariate with a random effect in addition to a column of ones coding for the random intercept.

Item bundle models with random dependency effects. An *item bundle* is a subset of items that share

some material (Rosenbaum, 1988). For instance, in a test of reading comprehension, several items may be based on the same reading paragraph in order to reduce the total test time. It is an unrealistic assumption that the responses on the items of the same item bundle are conditionally independent, given the ability underlying the test. However, the assumption that an item bundle is conditionally independent of other item bundles is still reasonable.

In this article, we discuss two approaches for incorporating item bundle effects into an IRT model. The first approach consists of defining an additional random effect for each item bundle. This is the approach developed by Bradlow, Wainer, and Wang (1999), also followed by Scott and Ip (2002). In the second approach, which was developed by Hoskens and De Boeck (1997), Jannarone (1986), and Kelderman (1984), item bundle effects are incorporated by introducing additional fixed effects. The second approach is outlined in the Polytomous Data section.

Note that there might be other causes of conditional dependence between responses besides the dependency that results from the use of item bundles. Another possibility is, for example, that the dependence results from learning during the test, as in the dynamic Rasch model, in which the dependency is also accounted for by fixed effects. The dynamic Rasch model is discussed in the *Person-by-Item Covariates* section.

We now discuss the approach proposed by Bradlow et al. (1999). The item bundle model of Bradlow et al. is a normal ogive model with an item discrimination parameter, but their approach can also be used for a logistic model without a discrimination parameter, as we do now. For this purpose, we extend the linear predictor ($\eta_{mi} = \theta_n + \beta_i$) of the Rasch model (see Equation 7) to

$$\eta_{mi} = \theta_n + \beta_i + \gamma_{nb(i)}, \quad (8)$$

where $\gamma_{nb(i)}$ is a random effect representing the person-specific effect for item bundle b that contains item i . The conditional dependency model by Bradlow et al. (1999) can be seen as a special case of the RWLLTM, in which additional item covariates with random effects are used to code for the item bundles.

When the model is extended with an item discrimination parameter, as in Bradlow et al. (1999) and in Scott and Ip (2002), it no longer belongs to the class of generalized linear mixed models. Such extended models are described in the Extending the Mixed Logistic Regression Model section.

Person Covariates: Latent Regression

When information is available about the participants taking a test, it might be interesting to estimate the extent to which the latent person parameter θ_n is determined by these person characteristics. This can be achieved by constructing a (multiple) regression model for the latent variable θ_n (Adams, Wilson, & Wu, 1997; Mislevy, 1987; Zwinderman, 1991):

$$\theta_n = \mathbf{x}'_n \mathbf{B} + \varepsilon_n, \quad (9)$$

where $\varepsilon_n \sim N(0, \sigma_\varepsilon^2)$ and \mathbf{x}_n is a vector of person covariates. In terms of the mixed logistic regression model, this means that \mathbf{X} now contains both person and item covariates. Furthermore, ε_n , the error term of the latent regression, which is normally distributed over persons, is now the random effect. Even if one is not interested in the relation between the latent variable and the person characteristics as such, including additional information about participants offers the advantages that more precise estimates for the (fixed) item and (random) person parameters are obtained (Mislevy, 1987). Furthermore, if the model of Equation 9 holds, then the unconditional distribution of θ_n is not normal but a mixture of normals (Verbeke & Molenberghs, 1997), invalidating the assumption of normally distributed random effects for a model without latent regression.

Similarly to the regression model for the random intercept θ_n , a regression model can be constructed for other random effects. For example, in the RWLLTM, the interaction between persons and an item covariate is accounted for by incorporating a random effect for that particular item covariate, rendering the effect of the item covariate person dependent. By regressing the random effect on person covariates, one can explain the person-dependent effect of the item covariate in terms of characteristics of the person.

Person-by-Item Covariates

Models for differential item functioning. In general, we can distinguish between two kinds of differences between groups in their performance on a test. First, it may be the case that groups differ with respect to the ability the test is measuring. This kind of difference can be modeled through a latent regression on the latent person variable or through multilevel modeling, as described in, respectively, the previous section and the *Multilevel IRT Models* section below.

A second scenario is that the groups under investigation do not differ with respect to ability but that one group is performing worse than the other(s) on

one or more of the items nevertheless. In other words, some of the items are biased against one group, or more generally, they are functioning differentially. An overview on differential item functioning can be found in Holland and Wainer (1993).

When an item is suspected of functioning differentially, this can be taken into account in the mixed logistic regression model by constructing person-by-item covariates as the product of person covariates coding for the different groups and an item covariate coding for the item.

Both the RWLLTM and models for differential item functioning deal with interactions between persons and items (item covariates), though in a different way. In the RWLLTM, the effect of an item covariate is allowed to be different for all persons. In contrast, in models for differential item functioning, the effect of an item can only differ between groups of persons. A second difference is that the RWLLTM incorporates random effects for item covariates to deal with interactions between persons and item covariates, whereas common models for differential item functioning include a fixed effect for each group by the construction of the person-by-item covariates just mentioned. The two differences are related: Including a fixed effect for each person would result in a multitude of parameters to be estimated.

Dynamic models. The dynamic Rasch model (Verhelst & Glas, 1993) is a model for learning and change during the test. In its most general form, the dynamic Rasch model reads as

$$p[Y_{ni} = 1 | \beta_i, \theta_n, f_i(\mathbf{y}^{ni}), k_i(\mathbf{r}^{ni})] = \frac{\exp[\theta_n + \beta_i + f_i(\mathbf{y}^{ni}) + k_i(\mathbf{r}^{ni})]}{1 + \exp[\theta_n + \beta_i + f_i(\mathbf{y}^{ni}) + k_i(\mathbf{r}^{ni})]} \quad (10)$$

or, equivalently,

$$L(\pi_{ni}) = \theta_n + \beta_i + f_i(\mathbf{y}^{ni}) + k_i(\mathbf{r}^{ni}), \quad (11)$$

where \mathbf{y}^{ni} is the partial response vector, consisting of participant n 's responses for items 1 to $i - 1$, and \mathbf{r}^{ni} is the partial reinforcement vector, consisting of reinforcements given after each of the responses up to item i . $f_i(\cdot)$ and $k_i(\cdot)$ are real-valued functions. Hence, the probability of giving a correct answer is a function of ability, of item difficulty, and of previous responses and reinforcements. That responses can be modeled as a function of other responses is an important model feature for the study of behavior, because behavior is a function not only of the person and the situation but also of previous behaviors.

In general, each dynamic Rasch model characterized by a specific choice for $f_i(\cdot)$ and $k_i(\cdot)$ corresponds to a mixed logistic regression model with accordingly defined person-by-item covariates because the possible function values of $f_i(\cdot)$ and $k_i(\cdot)$ can be considered as parameters. However, in its most general form, the dynamic Rasch model is not identified because the number of parameters outnumbers the number of possible response patterns (Verhelst & Glas, 1993, p. 397). Useful models do arise when some suitable restrictions are imposed on $f_i(\cdot)$ and $k_i(\cdot)$. We consider two examples.

In a first example, it is assumed that a constant amount of learning ζ takes place after each successfully solved item, but no reinforcement is given. Hence, $f_i(\mathbf{y}^{ni}) = \zeta \sum_{j=1}^{i-1} y_{nj}$, $i = 2, \dots, I$, and the term $k_i(\mathbf{r}^{ni})$ can be omitted. The model is fitted in the mixed logistic regression framework by including a person-by-item covariate that is defined as the number of items solved correctly by person n up to item i . Because the amount of learning is considered to be the same for all persons, the person-by-item covariate is only included in \mathbf{X} , the supermatrix containing the covariates with a fixed effect.

As a second example, consider the case in which solving a particular item i^* is facilitated if one succeeded in solving a related item j presented earlier in the test. Then, $f_i(\mathbf{y}^{ni}) = 0$ for all $i \neq i^*$, and $f_{i^*}(\mathbf{y}^{ni^*}) = \nu y_{nj}$, where ν represents the amount of facilitation. Again, the term $k_i(\mathbf{r}^{ni})$ can be omitted, because no reinforcement is taken into account. The facilitation effect can be incorporated by including a person-by-item covariate into the mixed logistic regression model that has, for item i^* , as its value the response on item j , y_{nj} , and is zero otherwise. Similar to the first example of the dynamic Rasch model, the person-by-item covariate is only included in \mathbf{X} because the amount of facilitation is considered to be the same for all persons.

The two examples considered for the dynamic Rasch model both involve the estimation of a learning parameter. By consequence, the corresponding parameters in the mixed logistic regression model are expected to be positive. In principle, negative dependencies among items may also occur, for example, because of interbehavioral compensatory or inhibitory mechanisms.

Multilevel IRT Models

Hitherto, we have considered only the case in which there is just one kind of clustering: measure-

ments clustered within participants. However, it might be that an additional clustering of the data is present. For example, consider the case in which participants are selected from several schools. Similar to the dependency between the responses given by a single participant, one might expect that the responses of participants belonging to the same school are not independent. More particularly, it might be unwarranted to assume that the latent person parameters θ_n or more generally the random effects, are independent realizations from a distribution defined over the population of participants.

Two strategies can be followed to account for additional cluster effects. First, if one is exclusively interested in the specific effects of the schools (or other clusters of participants), one can perform a latent regression with person covariates coding for the schools. This approach results in a separate fixed parameter for each school.

Alternatively, if the schools can be regarded as a random sample from a population of schools, and one wishes to draw conclusions with respect to this population, one can define the effects of the schools to be random, analogous to the random effects of participants. The latter approach results in multilevel IRT models (Fox & Glas, 2001; Kamata, 2001; Maier, 2001). Considering the standard case of a mixed logistic regression model with only a random intercept, and assuming normal distributions for the random effects on all levels, we obtain

$$L(\pi_{nsi}) = \mathbf{x}'_{nsi}\boldsymbol{\beta} + \theta_{ns}, \quad (12)$$

where $\theta_{ns} \sim N(\vartheta_s, \sigma_s^2)$, $\vartheta_s \sim N(0, v^2)$, with θ_{ns} denoting the random intercept for person n nested within school s , and ϑ_s denoting the random effect of school s . Hence, the joint distribution of the latent abilities θ_{ns} , $s = 1, \dots, S$, within a school is multivariate normal with mean $\mathbf{0}$, variances $\sigma_s^2 + v^2$, and covariances v^2 .

Polytomous Data

So far, we have considered only binary responses. However, psychometric data are often polytomous, for example, rating scale data, or cognitive tests for which one distinguishes between false, partially correct, and correct. Instead of dichotomizing the data, and applying the mixed logistic regression model for binary data, one can model the polytomous responses directly by extending the models for dichotomous data. This way, the maximal information that is contained in the data is preserved.

The mixed logistic regression model can handle polytomous responses by forming logits for polytomous data. Given $J_i + 1$ response categories ($0, \dots, J_i$), four different sets of J_i nonredundant logits can be formed (Agresti, 1989, 2002):

1. Baseline-category logits are the log odds of a category in question versus a baseline category. This approach leads to the nominal response model with a priori known discrimination parameters (Bock, 1972).
2. Adjacent-categories logits are the ordinary logits of the conditional probabilities of response j , given response j or $j + 1$ (or, alternatively, given response j or $j - 1$). This approach leads to the partial credit model (Masters, 1982) and the rating scale model (Andrich, 1978).
3. Continuation-ratio logits are the ordinary logits of the conditional probabilities of response j , given response j or higher (or, alternatively, given response j or lower). This approach leads to the sequential response model (Tutz, 1990).
4. Cumulative logits are the logits of cumulative probabilities (or, alternatively, the logits of the probability of a response of category j or higher). This approach leads to the graded response model with a priori known discrimination parameters (Samejima, 1969).

For binary data, the four types of logits are identical and correspond with the ordinary logits. The baseline-category logits are appropriate for nominal responses, whereas for ordinal responses, the three other types of logits are more appropriate because ordering information is taken into account. Furthermore, adjacent-categories logit models, but not continuation-ratio and cumulative logit models, can be expressed as baseline-category models (Agresti, 2002, chapter 7).

The mixed logistic regression model can now be formulated in logit terms as

$$L_{nij} = \mathbf{x}_{nij}\boldsymbol{\beta} + \mathbf{z}'_{nij}\boldsymbol{\theta}_n, \quad (13)$$

where L_{nij} is one of the logit transformations defined earlier and $j = 1, \dots, J_i$.

The supermatrices \mathbf{X} and \mathbf{Z} can be defined in a way analogous to the dichotomous case, but they now contain a separate row for each person-by-item-by-category combination, respectively, \mathbf{x}'_{nij} and \mathbf{z}'_{nij} (except for one category given that only J_i logits are nonredundant). Hence, if each item has the same num-

ber of response categories ($J_i = J$ for all i), and each person receives the same number of items, \mathbf{X} and \mathbf{Z} both consist of $N \times I \times J$ rows.

Again, we can distinguish between different kinds of covariates, depending on whether they are constant or not across items, persons, and/or categories. IRT models for polytomous data can be categorized according to two properties: the type of logit that is used as link function (Mellenbergh, 1995) and the kind of covariates they incorporate.

The rating scale model incorporates item covariates and category covariates. The other models for polytomous data (the nominal response, partial credit, sequential response, and graded response models) consider only item-by-category covariates. All models assume a unidimensional ability (only the intercept is random). Nevertheless, additional random effects can make sense and can easily be incorporated. Furthermore, similar to the latent regression models and the dynamic Rasch model for binary data, person covariates and/or person-by-item covariates could also be incorporated.

We discuss now two more models to illustrate the mixed logistic regression model for polytomous data. The first one is the partial credit model, often used to model ordinal response data. The second one illustrates the second approach of dealing with item bundle effects, which consists of introducing additional fixed parameters to account for the dependencies between the items of the same item bundle.

The Partial Credit Model

The *partial credit model* (Masters, 1982) incorporates the adjacent-category logit (Mellenbergh, 1995), contrasting category j with category $j - 1$. For an item with $J_i + 1$ categories ($0, \dots, J_i$), the partial credit model is defined as

$$L_{nij} = \log\left(\frac{\pi_{nij}}{\pi_{ni,j-1}}\right) = \theta_n + \beta_{ij}, \quad (14)$$

where π_{nij} is the conditional probability that the response of participant n on item i belongs to category j , $j = 1, \dots, J_i$. This results in the $J_i + 1$ category probabilities

$$\pi_{nij} = p(Y_{ni} = j | \boldsymbol{\beta}_i, \theta_n) = \frac{\exp\left[\sum_{v=1}^j (\theta_n + \beta_{iv})\right]}{1 + \sum_{h=1}^{J_i} \exp\left[\sum_{v=1}^h (\theta_n + \beta_{iv})\right]},$$

for $j > 0$, and

$$\pi_{ni0} = p(Y_{ni} = 0 | \boldsymbol{\beta}_i, \theta_n) = \frac{1}{1 + \sum_{h=1}^{J_i} \exp\left[\sum_{v=1}^h (\theta_n + \beta_{iv})\right]}, \quad (15)$$

where θ_n is a latent variable, and $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{iJ_i})'$. Equation 14 is a specification of Equation 13, where $\boldsymbol{\beta}$ is the vector resulting from stacking the I $\boldsymbol{\beta}_i$ vectors one below the other (with length $\sum_{i=1}^I J_i$); \mathbf{x}_{nij} has only one nonzero element, with value 1, to select parameter β_{ij} ; and \mathbf{z}_{nij} is 1 for all j, i , and n . Hence, \mathbf{X} consists of $\sum_{i=1}^I J_i$ item-by-category covariates, and \mathbf{Z} is a $N \times \sum_{i=1}^I J_i$ column vector of ones.

Item Bundle Models With Fixed Dependency Effects

Item bundle models with fixed dependency effects form part of the approach of Hoskens and De Boeck (1997) and were also described by Jannarone (1986) and Kelderman (1984). In these models, it is not an additional random effect that accounts for the conditional dependency between items of the same item bundle but a fixed parameter. For instance, the dependency between two items 1 and 2 is parameterized by an additional dependency parameter β_{12} . If items 1 and 2 make up a single item bundle, then the probability formula for a joint response on items 1 and 2 (with equal discrimination parameters of one) equals

$$p(Y_{n1} = y_{n1}, Y_{n2} = y_{n2} | \beta_1, \beta_2, \beta_{12}, \theta_n) = \frac{\exp[y_{n1}(\theta_n + \beta_1) + y_{n2}(\theta_n + \beta_2) + y_{n1}y_{n2}\beta_{12}]}{1 + \exp(\theta_n + \beta_1) + \exp(\theta_n + \beta_2) + \exp(2\theta_n + \beta_1 + \beta_2 + \beta_{12})}. \quad (16)$$

In Equation 16, the joint probability of a response pattern (y_{n1}, y_{n2}) is modeled. For an item bundle consisting of two binary items, there are four different response patterns possible: (0,0), (1,0), (0,1), and (1,1). Instead of looking at the two separate items as the basic units, one could also consider the item bundle as the basic unit and model the item bundle as if it were a polytomous item i^o with four different response categories 0, 1, 2, and 3 corresponding with the response patterns (0,0), (1,0), (0,1), and (1,1), respectively.

Taking the baseline-category logits of the newly defined item with 0 as baseline category, Equation 16 can be reformulated as follows:

$$L_{ni^{\circ}1} = \log\left(\frac{\pi_{ni^{\circ}1}}{\pi_{ni^{\circ}0}}\right) = \theta_n + \beta_1,$$

$$L_{ni^{\circ}2} = \log\left(\frac{\pi_{ni^{\circ}2}}{\pi_{ni^{\circ}0}}\right) = \theta_n + \beta_2, \quad (17)$$

and

$$L_{ni^{\circ}3} = \log\left(\frac{\pi_{ni^{\circ}3}}{\pi_{ni^{\circ}0}}\right) = 2\theta_n + \beta_1 + \beta_2 + \beta_{12}.$$

Equation 17 is a specification of Equation 13 where β equals $(\beta_1, \beta_2, \beta_{12})'$, $\mathbf{x}_{ni^{\circ}j}$ consists of two 0s and one 1 in appropriate places to select β_1 or β_2 if $j = 1$ or 2 , respectively, or consists of three 1s if $j = 3$. Furthermore, $z_{ni^{\circ}j}$ is equal to 1 for $j = 1, 2$ and equal to 2 if $j = 3$, and θ_n is the one-dimensional θ_n . An extension of this model to incorporate random dependency was described by Hoskens and De Boeck (1997) but is not considered here.

In summary, the key step to fit this type of dependency model in the framework of random effects models is to map the possible response patterns into distinct categories of a polytomous artificially constructed "item bundle item" i° , so that the correlational structure of the observations is modeled through one or more random effects (measuring the construct(s) underlying the test) and additional dependency parameters (to account for the dependency between the responses not accounted for by the random effects). For simplicity, we have assumed in our explanation that only two items from a test belong to an item bundle. The other items in the test may also be subdivided into additional item bundles (if they show conditional dependency) or they may not (if they are conditionally independent). The different item bundles in the test are assumed to be conditionally independent of each other. An item bundle may also consist of more than two dependent items. An application of such a model to projective testing can be found in Tuerlinckx, De Boeck, and Lens (2002).

Extending the Mixed Logistic Regression Model

In this section, we formulate an extension of the mixed logistic regression model. The extension consists of the inclusion into Equation 13 of loading or discrimination parameters that can be considered as item covariates with unknown values,

$$L_{nij} = \mathbf{x}'_{nij}\beta_1 + \mathbf{z}'_{nij}\theta_{n1} + \mathbf{a}'_i\beta_2 + \mathbf{b}'_i\theta_{n2}, \quad (18)$$

where \mathbf{x}_{nij} is a known P_1 -dimensional covariate vector for P_1 fixed effects, \mathbf{z}_{nij} is a known Q_1 -dimensional covariate vector for Q_1 random effects, \mathbf{a}_i is a P_2 -dimensional vector ($P_1 + P_2 = P$) for the unknown covariates of P_2 fixed effects, \mathbf{b}_i is a Q_2 -dimensional vector ($Q_1 + Q_2 = Q$) for the unknown covariates of Q_2 random effects, $\beta = (\beta'_1, \beta'_2)$ is the P -dimensional vector of fixed effects, and $\theta_n = (\theta'_{n1}, \theta'_{n2})$ is the Q -dimensional vector of random effects for unit n .

The extension in Equation 18 is formulated for polytomous data and hence also treats the case of binary data. The unknown covariates have no person index because we only allow unknown item covariates \mathbf{a}_i and \mathbf{b}_i . In the context of an ability test, the fact that an item covariate is unknown means that the degree to which a cognitive operation is involved in solving the item is not known in advance but is estimated from the data as a weight parameter similar to factor loadings.

Furthermore, we can define the supermatrices \mathbf{X} , \mathbf{Z} , \mathbf{A} , and \mathbf{B} in the same way as they were defined for the mixed logistic regression model. The supermatrices \mathbf{X} and \mathbf{Z} contain known covariate values. The supermatrices \mathbf{A} and \mathbf{B} contain unknown covariate values, or in other words, parameters. \mathbf{A} and \mathbf{B} consist of N identical submatrices, stacked one below the other because we only consider unknown item covariates.

The predictor (the right side of Equation 18) is no longer linear in that there appears a product of parameters, so that we are outside the family of generalized linear mixed models.

Again, we can distinguish between several types of covariates and categorize IRT models along these lines. Only IRT models that are specifications of Equation 18 and that take into account only unknown item covariates (and, for polytomous responses, unknown category covariates and item-by-category covariates) are formulated hitherto. For binary data, examples are the two-parameter logistic model (Birnbaum, 1968), the two-parameter logistic-constrained model (Embretson, 1999), the multidimensional two-parameter logistic model (McKinley & Reckase, 1983), the confirmatory multidimensional two-parameter logistic model (McKinley, 1989), and the model with internal restrictions on item difficulty (Butter, De Boeck, & Verhelst, 1998). The two-parameter model has also been presented with a probit link instead of a logit link (Bock & Aitkin, 1981; Lord & Novick, 1968).

For polytomous data, familiar IRT models of this type are the nominal response model in which the

discrimination parameters for the categories are restricted to be equal within items (Bock, 1972; baseline-category logits), the graded response model (Samejima, 1969; cumulative logits), and the generalized partial credit model (Muraki, 1992; adjacent-category logits). The unconstrained version of Bock's nominal response model allows that the discrimination parameters differ not only among items but also among the categories within a single item. The latter can be accommodated by allowing \mathbf{b}_i of Equation 18 to differ among categories and hence be subscripted with both an index i and j , \mathbf{b}_{ij} .

Analogously to the latent regression models and the dynamic Rasch model from the Mixed Logistic Regression Model for Binary Data section, it can be worthwhile to formulate IRT models that incorporate person covariates or person-by-item covariates. Another possibility is to regress the discrimination parameters on known item covariates, as is done in the two-parameter logistic-constrained model, or to restrict them otherwise, as is done by Thissen and Steinberg (1986). Thissen and Steinberg (1986) conceptualized a family of models that differ with respect to the restrictions placed on the discrimination parameters, ranging from the partial credit model (no discrimination parameters) to the nominal response model (no restrictions on the discrimination parameters). However, care should be taken in extending or adapting existing models not to make the model too complex to avoid identification problems. The general model of Equation 18 is illustrated in the following example.

According to the multidimensional two-parameter logistic model (McKinley & Reckase, 1983), the conditional probability of a correct answer is

$$\pi_{ni} = p(Y_{ni} = 1 | \beta_i, \boldsymbol{\theta}_n, \mathbf{b}_i) = \frac{\exp(\mathbf{b}_i' \boldsymbol{\theta}_n + \beta_i)}{1 + \exp(\mathbf{b}_i' \boldsymbol{\theta}_n + \beta_i)}. \quad (19)$$

When all the elements of \mathbf{b}_i are estimated, the model is exploratory. In the confirmatory model (McKinley, 1989), some elements of \mathbf{b}_i are constrained to zero. Hence, the unknown covariates collected in \mathbf{b}_i are only partially unknown in the latter case. Formulated in terms of the logit, the multidimensional two-parameter logistic model becomes

$$L_{ni} = \mathbf{b}_i' \boldsymbol{\theta}_n + \beta_i. \quad (20)$$

Equation 20 is a specification of Equation 18 for binary data, where each \mathbf{X}_n is an $I \times I$ identity matrix (and therefore the vector \mathbf{x} can be omitted, as was also the case for the Rasch model; see Equation 6), and θ_{n2}

equals θ_{n1} . The two other terms of Equation 18, $\mathbf{z}'_{nij} \boldsymbol{\theta}_{n1}$ and $\mathbf{a}'_i \boldsymbol{\beta}_2$, do not appear in Equation 20.

In the exploratory model, one restriction for each random effect or dimension is needed to render the model identifiable, because one can multiply a particular element of \mathbf{b}_i by a constant and divide the corresponding element of $\boldsymbol{\theta}_n$ by the same constant. The usual restrictions consist of fixing all the variances to one. An additional unidentifiability results from the rotational freedom with respect to $\boldsymbol{\theta}_n$, similar to the rotational freedom in classical factor analysis (Bock, Gibbons, & Muraki, 1988). The latter can be solved by fixing the correlations between the $\boldsymbol{\theta}_n$ and/or by putting some constraints on the \mathbf{b}_i . Alternatively, one can rotate the solution according to some criterion, such as varimax (orthogonal dimensions) or oblimin (correlated dimensions).

Other IRT Models

The models presented in Equations 5, 13, or 18 are quite general and capture the majority of commonly used IRT models. We now consider some categories of IRT models that do not fit within this framework. A first category consists of IRT models that incorporate both qualitative and quantitative latent person variables. In these models, the qualitative latent variables denote the latent class to which a person belongs. Within a class, an IRT model incorporating quantitative latent variables is assumed to hold. Hence, the first category of models consists of models that are a discrete mixture of other IRT models. Examples of this category are as follows: the hybrid model of Yamamoto (1987), which is a discrete mixture of IRT and random guessers; the mixed Rasch model (Rost, 1990; *mixed* referring to the discrete mixture, and not to the random intercept within each discrete mixture component); and the model of Mislevy and Verhelst (1990), which consists of a discrete mixture of LLTMs.

A second category consists of IRT models that include guessing parameters, such as the three-parameter logistic model (Birnbaum, 1968) for binary data. For polytomous data, see in this respect Samejima (1969) and Thissen and Steinberg (1984). These models can also be seen as discrete mixtures: On a particular item, a participant either chooses intentionally for a certain response alternative or "guesses" for a certain response alternative when "undecided" (Thissen & Steinberg, 1984, p. 502). In this second category of IRT models, the mixture is defined over

all measurements: Whether a participant is in the ignorance state or not may change from item to item. In contrast, for the first category of models, the mixture is defined over participants: They are in the same class for all items.

Throughout the article, the item parameters were considered fixed. Consequently, IRT models in which the effects of the items are random over items form a third category of models not captured by our framework (Janssen & De Boeck, 2003).

Statistical Inference and Software

In this section, we give an overview of estimating the mixed models considered in this article and discuss some software that can be used for estimating these models. In a final paragraph, we briefly mention some issues on model evaluation.

The following marginal likelihood has to be maximized to obtain parameter estimates:

$$L(\boldsymbol{\delta}, \boldsymbol{\Sigma} | \mathbf{y}_1, \dots, \mathbf{y}_N) = \prod_{n=1}^N L_n(\boldsymbol{\delta}, \boldsymbol{\Sigma}) \\ = \prod_{n=1}^N \int p(\mathbf{y}_n | \boldsymbol{\delta}, \boldsymbol{\theta}_n) \mathbf{N}(\boldsymbol{\theta}_n | \mathbf{0}, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_n, \quad (21)$$

where $p(\mathbf{y}_n | \boldsymbol{\delta}, \boldsymbol{\theta}_n)$ is the probability of observing response pattern \mathbf{y}_n and $\mathbf{N}(\boldsymbol{\theta}_n | \mathbf{0}, \boldsymbol{\Sigma})$ is the multivariate normal distribution (of dimension Q) with mean vector zero and covariance matrix $\boldsymbol{\Sigma}$. The parameter vector $\boldsymbol{\delta}$ contains the fixed effects that appear in $p(\mathbf{y}_n | \boldsymbol{\delta}, \boldsymbol{\theta}_n)$. If the model is a generalized linear mixed model, the parameter vector $\boldsymbol{\delta}$ equals $\boldsymbol{\beta}$, as in Equation 2.

For all mixed models considered in this article, the integral in Equation 21 is intractable. There are three general types of solutions to this problem. The first one is to approximate the integral with numerical integration techniques; this is called a *full-likelihood analysis*. The second solution consists of approximating the nonlinear model by a linear one and then estimating the parameters using an estimation algorithm for linear mixed models. Both solutions are developed in a frequentist context. As a third option, we mention Bayesian estimation methods.

Full-Likelihood Analysis

In a full-likelihood analysis, a numerical approximation to the likelihood in Equation 21 is maximized. Roughly speaking, there are four different possible

approaches, and they can be classified in a two-by-two table. The first dimension of the table classifies the techniques according to whether the numerical approximation to the marginal likelihood is maximized directly or whether the maximization problem is transferred to another function for which it can be shown that as a by-product, the marginal likelihood is maximized, too. The second dimension distinguishes between deterministic and stochastic numerical approximations to intractable integrals encountered in the optimization problem.

Direct maximization. In direct maximization techniques, it is the intractable integral in Equation 21 that is numerically approximated and then this numerical approximation is maximized. A first possibility is to approximate the integral numerically in a deterministic way by means of a numerical quadrature rule. In the case in which the random effects are assumed to be normally distributed, usually the Gauss–Hermite quadrature is chosen. A second possibility is to use Monte Carlo integration (i.e., the stochastic option).

Currently there are four popular software packages that provide a full-likelihood analysis with direct optimization for at least some IRT models: SAS PROC NLMIXED, GLLAMM (generalized linear latent and mixed models; Rabe-Hesketh, Pickles, & Skrondal, 2001; Skrondal & Rabe-Hesketh, in press), MIXOR (mixed effect ordinal regression; Hedeker & Gibbons, 1996), MIXNO (mixed effect nominal logistic regression; Hedeker, 1999), and HLM (hierarchical linear models; Version 5; Raudenbush, Bryk, & Congdon, 2001). All programs allow the use of person, item, and person-by-item covariates.

Almost all IRT models discussed in this article (both generalized linear and nonlinear mixed models) can be estimated using PROC NLMIXED in SAS; the only exceptions are multilevel models with three or more levels. The intractable integral can be approximated by a Gauss–Hermite quadrature, an adaptive Gauss–Hermite quadrature, or their stochastic counterparts (Pinheiro & Bates, 1995). Moreover, there are a variety of methods available to maximize the marginal loglikelihood that differ from each other in the order of the derivatives that are used.

GLLAMM is a program written for STATA (Stata-Corp., 2001), and it allows one to fit all models discussed in this article, including models with more than two levels and models with a discrimination parameter. To approximate the integral in Equation 21, the user may choose between a regular or an adaptive

Gauss–Hermite quadrature. One can also choose a semiparametric estimation method, in which the mixing distribution is discrete and the node points and corresponding weights are estimated. The latter method is appropriate when the assumption of a normal mixing distribution is unwarranted. The maximization is done using a Newton–Raphson algorithm.

The programs MIXOR and MIXNO use a Gauss–Hermite quadrature to approximate the intractable integral. MIXOR can be used for estimating IRT models for dichotomous data that are generalized linear mixed models. Also the two-parameter logistic model can be estimated. As for polytomous data, MIXOR is suited for the graded response models with equal discrimination. MIXOR can handle only two-level data. MIXNO can be used to estimate not only multinomial logit models with and without discrimination parameters but also more restricted versions of these models, such as the partial credit model and dependency models with fixed effects for the item bundles. Both in MIXOR and MIXNO, a Fisher scoring algorithm is used to maximize the marginal likelihood.

In the newest version of HLM (Version 5), a sixth-order Laplace approximation to the intractable integral is implemented (Raudenbush, Yang, & Yosef, 2000). Although different from the adaptive Gaussian quadrature (Pinheiro & Bates, 1995), the Laplace approximation is related to it, and therefore it is listed under the full-likelihood analysis. Currently, it is implemented only for the regular Rasch-type models (generalized linear mixed models with only two response categories and Bernoulli distributed data) with a maximum of two levels. Other types of generalized linear mixed model IRT models can be estimated by HLM but only by means of a linearized analytical approximation (see below).

Indirect maximization. Indirect maximization in mixed effects models is based on the Expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977; for other types of indirect maximization, see Lange, Hunter, & Yang, 2000). The random effects are considered as missing data, and together with the observed data they form the complete data. Because the random effects are not known, one first computes, given the current estimates of the fixed effects and observed data, the expected value of the complete loglikelihood (the E-step), and then expected loglikelihood is maximized (the M-step). In the E-step, a similar intractable integral as the one in Equation 21 appears, and it has to be approximated again using the aforementioned methods.

As noted by Bock and Aitkin (1981), for the standard versions of the IRT models (hence without additional covariates) the application of the EM algorithm has a major advantage. In that case, the item parameter vector δ can be subdivided into I subsets of parameters, each pertaining to only one item. Consequently, the expected loglikelihood can be written as a sum of independent terms that can be maximized separately because, given the random effects, there is independence between the items. This approach allows one to analyze a large number of items. The EM algorithm is currently implemented in IRT software packages such as MULTILOG (Thissen, 1991) and ConQuest (Wu et al., 1998).

MULTILOG can handle Rasch models, two-parameter logistic models, graded response models, and multinomial logit models (the last two with and without discrimination parameters). Also restricted versions of the multinomial logit model such as the partial credit model and fixed effects dependency models can be estimated. MULTILOG can handle multiple groups, which is a specific case of latent regression. The intractable integral in Equation 21 is approximated by a Gauss–Hermite quadrature. One can also opt for a semiparametric method when the assumption of a normal mixing distribution is unwarranted.

ConQuest can handle all IRT models of the generalized linear mixed model type that incorporate person and/or item covariates. Also models for differential item functioning can be estimated. The intractable integral in Equation 21 is approximated by Gauss–Hermite quadrature or its stochastic counterpart.

Spiessens, Verbeke, and Komárek (2002; see Spiessens, Verbeke, Fieuws, & Rijmen, 2003, for an application) proposed an EM algorithm implemented in a SAS macro that can be used to approximate a nonnormal mixing distribution by a discrete mixture of normal distributions. In this case, the missing data consist of the respective component of the discrete mixture to which persons belong. The M-step is carried out with PROC NL MIXED.

Linearized Analytical Approximations

The integral in Equation 21 has a closed-form solution if the distribution of the data conditional on the random effects is normal and the link function is the identity function. In that case, the marginal distribution is also normal (see Verbeke & Molenberghs, 2000). Because the estimation methods for the normal linear mixed model are well established, many re-

searchers have tried to apply them also in the case of generalized linear and nonlinear mixed models through approximating the nonlinear model with a linear one. Fortunately, the assumption of normality is not necessary to estimate the parameters because only the mean and variance–covariance matrix of the data have to be specified; this is called the *quasi-likelihood approach* (McCullagh & Nelder, 1989).

Therefore, a strategy to avoid the intractable integral is to use a linear approximation (by means of a first-order Taylor expansion) to the link function. In a next step, the mean and variance–covariance matrix for the data can be derived, and then one of the linear mixed model estimation methods can be applied. These two steps (linearization and estimation) are repeated until convergence.

There are two popular approaches. The first one is the penalized quasi-likelihood method (PQL; Breslow & Clayton, 1993) in which the Taylor expansion of the link function is around the current estimates for the fixed effects, and around the empirical Bayes estimates for the random effects. The second approach is the marginal quasi-likelihood method (MQL; Goldstein, 1991) in which the Taylor expansion of the link function is around the current estimates for the fixed effects and zero for the random effects. Several improved extensions of both methods have been proposed (MQL2, PQL2, and corrected PQL; see Breslow & Lin, 1995; Goldstein & Rasbash, 1996; Rodriguez & Goldman, 1995). The quasi-likelihood methods have been investigated only for generalized linear mixed models and for nonlinear mixed models with normally distributed error (Wolfinger & Lin, 1997); hence, it is unclear whether they can be applied to models such as the two-parameter logistic model.

Rodriguez and Goldman (1995) showed in a simulation study that fixed and/or variance components estimated with MQL may suffer from a downward bias, especially when random effects are large and the number of subunits for a unit is small, and Breslow and Clayton (1993) showed that such biases also occur for PQL. For the unidimensional Rasch model for two items, Breslow and Lin (1995) showed that the PQL estimates for the regression coefficients also have an appreciable downward asymptotic bias; that led these authors to propose a corrected version of the PQL (not equal to PQL2) by adding a term to overcome the asymptotic bias. Also the variance components have some asymptotic bias when estimated with PQL, and here a multiplicative correction factor is proposed. Lin and Breslow (1996) extended the re-

sults from Breslow and Lin (1995) to generalized linear mixed models with multiple independent random effects.

Both the MQL and PQL methods are implemented in MLwiN (Rasbash, Browne, Goldstein, & Yang, 2000) to estimate generalized linear mixed models for binary data (currently it is not possible to estimate models for polytomous data). The PQL method is implemented in HLM (Version 5), and multilevel generalized linear mixed models up to three levels can be estimated in HLM (Version 5) for binary data. For polytomous data, only models with at most two levels can be estimated.

Bayesian Methods

An important feature of the aforementioned estimation methods is that they are developed within a frequentist framework of statistical inference. An alternative is to consider Bayesian estimation methods. Modern computer-intensive techniques known as *Markov chain Monte Carlo methods* (Gelman, Carlin, Stern, & Rubin, 1995; Tanner, 1996) often make the parameter estimation problem in a Bayesian framework less complex than techniques in a frequentist one, but the drawback is that the optimization techniques are often more time-consuming. Currently, Bayesian estimation can be done using BUGS (Gilks, Thomas, & Spiegelhalter, 1994) and MLwiN.

Model Evaluation

Model evaluation is used as a collective term for activities such as hypothesis testing, constructing confidence intervals, model selection, variable selection, graphical model checking, and so on. Inferential procedures that are applicable for fixed effects generalized linear and nonlinear models can usually be used for inferences about fixed effects in the mixed models without many problems (Wald tests, likelihood ratio tests, etc.). However, for inferences about variance components, some caution is necessary. For instance, the reference distributions of the likelihood ratio test statistic require some modification to test whether variance components differ from zero (see Self & Liang, 1987; Stram & Lee, 1994). Also, it has to be stressed that the deviance statistic produced by using PQL or MQL methods (or their extensions) cannot be used in the subsequent model stage because they lack the necessary asymptotic properties (Snijders & Bosker, 1999). In our opinion, none of the current software packages offer satisfactory possibilities for model evaluation.

Application: A Self-Report Study on Anger

As an example, we discuss a questionnaire in which hypothetical aversive situations were judged on the degree to which they elicit anger (Kuppens, 2003). The questionnaire consisted of 24 hypothetical situations, such as “You’re at a party when you hear that someone broke your bike,” and “One day, a member of your family is ill with 39.5 °C (103 °F) of fever.” The English translation (translated from Dutch) of the list of 24 situations used can be obtained from Peter Kuppens. The questionnaire was presented to 510 high school students, 179 (35%) female and 331 (65%) male, in groups of about 20 students. The mean age of the students was 17 years ($SD = 1$). The participants rated on a 4-point scale (ranging from 0–3) to what extent each of the situations elicited anger. Additional information obtained for the participants consisted of the mean score on the 10 items of the trait anger scale of Spielberger (Van der Ploeg, Defares, & Spielberger, 1982; ANGER), the mean score on the 10 items of the Rosenberg Self-Esteem Scale (Rosenberg, 1989; SELF), the mean score on a subset of 10 items of the irritation scale of Eysenck (Eysenck, 1953; IRRITATION), and the gender of the participants (GENDER). For ANGER, SELF, and IRRITATION, the mean item scores (items were all scored from 0–3) were 1.3, 1.8, and 1.8, respectively. The standard deviations were, respectively, 0.6, 0.6, and 0.5. The intercorrelations were rather low, with the highest correlation of .31 being between IRRITATION and ANGER ($p < .001$).

A second group of 25 first-year psychology undergraduate students, 19 (76%) female and 6 (24%) male, with a mean age of 19 years ($SD = 2$), rated the 24 situations on a 4-point scale (also ranging from 0–3) with respect to a set of 10 situational characteristics. The students participated in the study as a partial fulfillment of their course credits. The situational characteristics to judge were the amount of control over the situation (CONTROL), whether the situation could be changed (CHANGE), the predictability of the situation (PREDICT), the consequences for a third person (CONSEQ3), the consequences for oneself (CONSEQ1), whether the situation was threatening for oneself (THREAT), whether one had to blame oneself (BLAME), whether the situation touched oneself in a personal emotional way (TOUCH), whether what was going on was clear (CLEAR), and whether one had a loss experience in the situation (LOSS).

We report on two illustrative analyses. A more profound set of analyses would be beyond the aim of this

article. The first analysis is targeted on the prediction of the anger responses on the basis of person and item characteristics. The second is targeted on revealing the underlying structure of the questionnaire.

Rating Scale Model With a Decomposition of the Item Location Parameters and Latent Regression

The base model for the first analysis is the rating scale model (Andrich, 1978), which is a partial credit model (see Equations 14 and 15) in which the item parameters are decomposed as $\beta_{ij} = \beta_i + \delta_j$, where β_i is the location of item i , and δ_j is the deviation of category j from the item location. The latter are assumed equal across items, so that the only difference among items is the difference in location. To identify the model, we set δ_1 to zero.

Furthermore, the item location parameters β_i were decomposed into a weighted sum of item covariates, analogous to the LLTM. As item predictors, we used the mean ratings, computed over participants, of the situations on the 10 situational characteristics.

Finally, the latent variable of the rating scale model, θ_n , was modeled through a latent regression (see Equation 9), with ANGER, SELF, IRRITATION, and GENDER serving as person covariates.

Putting all those components together, the full model is, for $j = 1, 2, 3$ because there were four response categories,

$$\log\left(\frac{\pi_{nij}}{\pi_{ni,j-1}}\right) = \varepsilon_n + \beta_{anger} \text{ANGER}_n + \beta_{self} \text{SELF}_n + \beta_{irritation} \text{IRRITATION}_n + \beta_{gender} \text{GENDER}_n + \beta_{control} \text{CONTROL}_i + \beta_{change} \text{CHANGE}_i + \beta_{predict} \text{PREDICT}_i + \beta_{conseq3} \text{CONSEQ3}_i + \beta_{conseq1} \text{CONSEQ1}_i + \beta_{threat} \text{THREAT}_i + \beta_{blame} \text{BLAME}_i + \beta_{touch} \text{TOUCH}_i + \beta_{cope} \text{CLEAR}_i + \beta_{loss} \text{LOSS}_i + \delta_j. \quad (22)$$

The first 5 terms define the latent regression of the latent variable on person covariates (hence the participant subscript n), whereas the next 10 define the decomposition of the item location into a weighted sum of item covariates (hence the item subscript i). Equation 22 is a mixed logistic regression model for polytomous data; see Equation 13. \mathbf{Z} is a column vector of ones of length $510 \times 24 \times 3 = 36,720$. \mathbf{X} consists of the same number of rows and 17 columns:

a column vector of ones (intercept), four person covariates (ANGER to GENDER) defining the latent regression, 10 item covariates (CONTROL to LOSS) defining the decomposition of the item parameters, and two category covariates to select the appropriate δ_j (only two, because setting δ_1 to zero is equivalent to removing the corresponding covariate from \mathbf{X}).

The model was estimated using the NLMIXED procedure of SAS. The program code is presented with some discussion in Appendix A, which is available in the online version of this article in the PsycARTICLES database. The parameter estimates, standard errors, and significance probabilities for the t statistic with 509 degrees of freedom¹ are shown in Table 1.

With respect to the decomposition of the item parameters into a weighted sum of the situational characteristics, all of them do contribute to the prediction (all $ps < .05$) despite the high intercorrelations between several of them. The situational characteristics with positive regression weights were, from high to low, CHANGE, LOSS, CLEAR, PREDICT, TOUCH, and THREAT, with regression parameters of, respectively, 1.82, 1.46, 1.02, 0.21, 0.17, and 0.13. CONSEQ3, BLAME, CONSEQ1, and CONTROL all

had negative regression weights of, respectively, -0.11 , -0.21 , -0.45 , and -1.93 . As in ordinary multiple regression analysis, the regression weights represent the contribution of a particular covariate to the prediction of anger, given all other covariates. Especially when intercorrelations between covariates are high, the estimate of a regression weight may heavily depend on which other covariates were included in the model. Some of the intercorrelations between the item covariates were quite high indeed, up to .86 between CONTROL and CHANGE ($p < .01$). Removing the item covariates having high correlations with other item covariates might render the interpretation more clear. However, because all regression weights were significant, removing some of the item covariates would result in a worse fit of the model to the data.

With respect to the latent regression parameters, ANGER and IRRITATION have positive regression weights of 0.32 and 0.39, respectively (both $ps < .01$); the weight of SELF is not significantly different from zero ($\beta_{self} = 0.05$, $p = .28$); and GENDER has a marginally significant negative regression weight of -0.10 ($p = .05$). The variance of the latent variable not explained through the latent regression, $\sigma_{\epsilon_n}^2$, amounted to 0.24. Hence, given the other covariates, participants with a high score on the anger scale tended to exhibit more anger, participants with a high score on the irritation scale also tended to exhibit more anger, the score on the self-esteem scale was not predictive, and male participants were more likely than female participants to exhibit anger (female participants were coded with a 1 on GENDER, and male participants with a 0). Anger being the dependent variable, it was somewhat surprising that the regression weight of IRRITATION was slightly higher than the regression weight of ANGER. The latter was not due to the presence of more variation in IRRITATION, the standard deviation of IRRITATION being smaller than the standard deviation of ANGER,

Table 1
Parameter Estimates, Standard Errors, and Significance Probabilities of the Rating Scale Model With Latent Regression and Decomposition of the Item Location Parameters

Parameter	Estimate	SE	p ($df = 509$) ^a
β_{anger}	0.32	0.05	<.01
β_{self}	0.05	0.04	.28
$\beta_{irritation}$	0.39	0.06	<.01
β_{gender}	-0.10	0.05	.05
$\beta_{control}$	-1.93	0.07	<.01
β_{change}	1.82	0.06	<.01
$\beta_{predict}$	0.21	0.06	<.01
$\beta_{conseq3}$	-0.11	0.03	<.01
$\beta_{conseq1}$	-0.45	0.06	<.01
β_{threat}	0.13	0.06	.03
β_{blame}	-0.21	0.03	<.01
β_{touch}	0.17	0.06	<.01
β_{clear}	1.02	0.09	<.01
β_{loss}	1.46	0.06	<.01
δ_2	-0.31	0.05	<.01
δ_3	-1.02	0.05	<.01
Intercept	-4.48	0.20	<.01
$\sigma_{\epsilon_n}^2$	0.24	0.02	<.01 ^b

^a Probability values for t statistic. ^b Significance probabilities for the variance of the error term of the latent regression should not be relied upon (see the Statistical Inference and Software section).

¹ For single effects, it is common to use the standard normal distribution as the reference distribution for the t statistic (the parameter estimate divided by its estimated standard error); this is known as the Wald test (Fahrmeir & Tutz, 2001). In contrast, SAS NLMIXED uses the Student's t distribution as the reference distribution for the t statistic, with the number of degrees of freedom being equal to the number of units minus the number of random effects. We adhere to the output as given by SAS NLMIXED. However, with 509 degrees of freedom, the Student's t distribution is almost identical to the standard normal distribution.

or a higher reliability of the irritation scale, Cronbach's alpha being lower for the irritation scale than for the anger scale, with values of .70 and .85, respectively.

Finally, the estimates of the category parameters δ_2 and δ_3 were -0.31 and -1.02 , respectively (both $ps < .01$). We note that the ordering of the category parameters, from 0 for δ_1 to -1.02 for δ_3 , is an empirical result and is not imposed by the model. Such an ordering of category parameters would have been imposed if cumulative logits were used, however. The ordering of the category parameters implies $-\beta_{i1} < -\beta_{i2} < -\beta_{i3}$ for all i , because the item parameters were decomposed as $\beta_{ij} = \beta_i + \delta_j$. Because $-\beta_{ij}$ represents the value of θ_n where the probability curves for category $j - 1$ and j intersect (Masters, 1982, p. 162), the ordering of the category parameters means that for each category, a region of θ_n existed for which that particular category is the most likely to occur.

Two-Dimensional Two-Parameter Logistic Model

For the second analysis, the responses were dichotomized: zero and one were recoded as a zero, and two and three were recoded as a one. These dichotomized responses were analyzed with an exploratory two-dimensional two-parameter logistic model (McKinley & Reckase, 1983; see Equations 19 and 20). To identify the model, we set the variances of the latent variables to one; we set the covariance between the two latent variables to zero (thus the latent variables were constrained to be orthogonal); and finally we also set b_{12} , the value of the first item on the second unknown item covariate, to one.

The exploratory two-dimensional two-parameter logistic model considered here is a particular instance of the nonlinear mixed model of Equation 18. \mathbf{X} consists of 510 identity matrices of size 24×24 , stacked one below the other. \mathbf{B} consists of 510 identical matrices of size 24×2 , with i th row the vector (b_{i1}, b_{i2}) , $i = 1, \dots, 24$, also stacked one below the other. \mathbf{Z} and \mathbf{A} are not defined; see the discussion of the multidimensional, two-parameter logistic model.

The model was estimated using the NLMIXED procedure of SAS. The program code is presented with some discussion in Appendix B, which is available in the online version of this article in the PsycARTICLES database. The parameter estimates, standard errors, and significance probabilities for the t statistic with 508 degrees of freedom are given in Table 2.

Revealing the underlying structure of the questionnaire as the main target of the second analysis, we focused on the estimates of the unknown item covariates. In Figure 1, the estimates of the unknown item covariates for the first random effect (the loadings of the items on the first dimension) are plotted against the estimates of the unknown item covariates for the second random effect (the loadings of the items on the second dimension).

Two groups of items can be distinguished in Figure 1: one group with relatively low loadings on the first dimension and moderate to high loadings on the second dimension, and a second group with relatively high loadings on the first dimension and low to moderate loadings on the second dimension. The first group consists of WAITER, SOMEONE ELSE, COMPETITION, ILL, GLASSES, CROSSING, HOME, SAME FEELINGS, SWIM, COMA, and TROUSERS. The second group consists of SON, LOSS, DISK, MESS, JOBSTUDENT, OLD MAN, BORROW, HOMEWORK, NOISE, NOBODY HOME, BIKE, and GOSSIP.

According to descriptions of the situations, the second group of items turned out to consist of situations in which the aversive situation is caused by a third person, either because the latter purposely behaved or did not behave on purpose in a specific way. In the first group, to the contrary, no such third person is present. Only DISK ("You have to write a thesis for school. You write it on the computer. A week before the deadline, your disk gets jammed into the computer. This causes your disk to break, and all your work is lost.") does not fit within this interpretation, as it belongs to the second group and there is no third person present. However, one can consider the computer to play the role of being an entity that refuses to function properly in this situation.

Concluding Remarks

We proposed a nonlinear mixed model framework for IRT models, relating psychometrics to a broad statistical literature. Many IRT models, among them the "standard" IRT models, nicely fit into the nonlinear mixed model framework. Casting different models into this framework is a way of making explicit the differences and commonalities among IRT models. Furthermore, standard IRT models can readily be adapted and extended, and the resulting models can be estimated using existing software for generalized linear and nonlinear mixed models.

Table 2
Parameter Estimates, Standard Errors, and Significance Probabilities of the Two-Dimensional Two-Parameter Logistic Model

Parameter	Estimate	SE	<i>p</i> (<i>df</i> = 508) ^a	Parameter	Estimate	SE	<i>p</i> (<i>df</i> = 508) ^a
β_{waiter}	-1.92	0.15	<.01	$b_{homework,1}$	1.47	0.24	<.01
$\beta_{someone\ else}$	-1.40	0.15	<.01	$b_{same\ feelings,1}$	0.35	0.24	.14
$\beta_{competition}$	-1.93	0.16	<.01	$b_{noise,1}$	1.14	0.21	<.01
β_{ill}	-4.61	0.49	<.01	$b_{nobody\ home,1}$	1.27	0.21	<.01
β_{son}	2.96	0.31	<.01	$b_{test,1}$	0.39	0.26	.14
β_{loss}	1.79	0.18	<.01	$b_{swim,1}$	0.26	0.23	.27
$\beta_{glasses}$	-1.41	0.17	<.01	$b_{bike,1}$	1.41	0.27	<.01
β_{disk}	2.62	0.25	<.01	$b_{gossip,1}$	1.58	0.29	<.01
β_{mess}	1.58	0.18	<.01	$b_{coma,1}$	0.46	0.19	.02
$\beta_{jobstudent}$	0.92	0.15	<.01	$b_{trousers,1}$	0.29	0.25	.25
$\beta_{crossing}$	-0.29	0.11	<.01	$b_{waiter,2}$ ^b	1	/	/
β_{home}	-3.55	0.40	<.01	$b_{someone\ else,2}$	1.07	0.21	<.01
$\beta_{old\ man}$	-0.68	0.12	<.01	$b_{competition,2}$	0.79	0.20	<.01
β_{borrow}	2.48	0.25	<.01	$b_{ill,2}$	0.49	0.55	.37
$\beta_{homework}$	0.61	0.14	<.01	$b_{son,2}$	0.33	0.31	.29
$\beta_{same\ feelings}$	-1.61	0.18	<.01	$b_{loss,2}$	0.52	0.25	.03
β_{noise}	0.37	0.13	<.01	$b_{glasses,2}$	1.37	0.28	<.01
$\beta_{nobody\ home}$	0.95	0.14	<.01	$b_{disk,2}$	0.45	0.29	.12
β_{test}	-2.61	0.25	<.01	$b_{mess,2}$	0.22	0.25	.38
β_{swim}	-2.28	0.21	<.01	$b_{jobstudent,2}$	0.73	0.27	<.01
β_{bike}	2.41	0.23	<.01	$b_{crossing,2}$	0.79	0.18	<.01
β_{gossip}	2.73	0.27	<.01	$b_{home,2}$	1.60	0.36	<.01
β_{coma}	0.39	0.11	<.01	$b_{old\ man,2}$	0.55	0.21	<.01
$\beta_{trousers}$	-1.90	0.20	<.01	$b_{borrow,2}$	0.78	0.32	.01
$b_{waiter,1}$	0.57	0.21	<.01	$b_{homework,2}$	0.84	0.25	<.01
$b_{someone\ else,1}$	0.26	0.20	.21	$b_{same\ feelings,2}$	1.37	0.27	<.01
$b_{competition,1}$	0.12	0.20	.55	$b_{noise,2}$	0.97	0.22	<.01
$b_{ill,1}$	0.35	0.57	.54	$b_{nobodyhome,2}$	0.62	0.23	<.01
$b_{son,1}$	1.60	0.31	<.01	$b_{test,2}$	1.28	0.27	<.01
$b_{loss,1}$	1.29	0.23	<.01	$b_{swim,2}$	1.13	0.25	<.01
$b_{glasses,1}$	0.62	0.25	.01	$b_{bike,2}$	0.40	0.29	.16
$b_{disk,1}$	1.51	0.28	<.01	$b_{gossip,2}$	0.52	0.30	.09
$b_{mess,1}$	1.46	0.25	<.01	$b_{coma,2}$	0.94	0.21	<.01
$b_{jobstudent,1}$	1.62	0.27	<.01	$b_{trousers,2}$	1.31	0.27	<.01
$b_{crossing,1}$	0.59	0.17	<.01	$\sigma_{\theta_{n1}^2}^b$	1	/	/
$b_{home,1}$	0.27	0.32	.40	$\sigma_{\theta_{n2}^2}^b$	1	/	/
$b_{old\ man,1}$	1.13	0.20	<.01	$\sigma_{\theta_{n1}\theta_{n2}}^b$	0	/	/
$b_{borrow,1}$	1.67	0.31	<.01				

^a Significance probabilities for *t* statistic. ^bParameters set to the indicated value to identify the model. Slashes indicate that no standard errors and significance probabilities can be computed.

Within the nonlinear mixed model framework, existing IRT models can be categorized according to several principles. A first principle, reflected in the structure of the article, is whether an IRT model is a mixed logistic regression model, is a member of the set of extended mixed logistic regression models, or does not fit within either set of models. Second, models can be categorized with respect to what kind of covariates they incorporate (item covariates, person covariates, etc.). Third, a distinction can be made be-

tween models for dichotomous and models for polytomous responses. With respect to the latter, one can make a further distinction based on the type of generalized logit used. Fourth, unidimensional models (one random effect, usually the intercept) can be contrasted with multidimensional models (more than one random effect).

Crossing the four categorization principles (which are not intended to form an exhaustive set of principles) would reveal a lot of new models to be for-

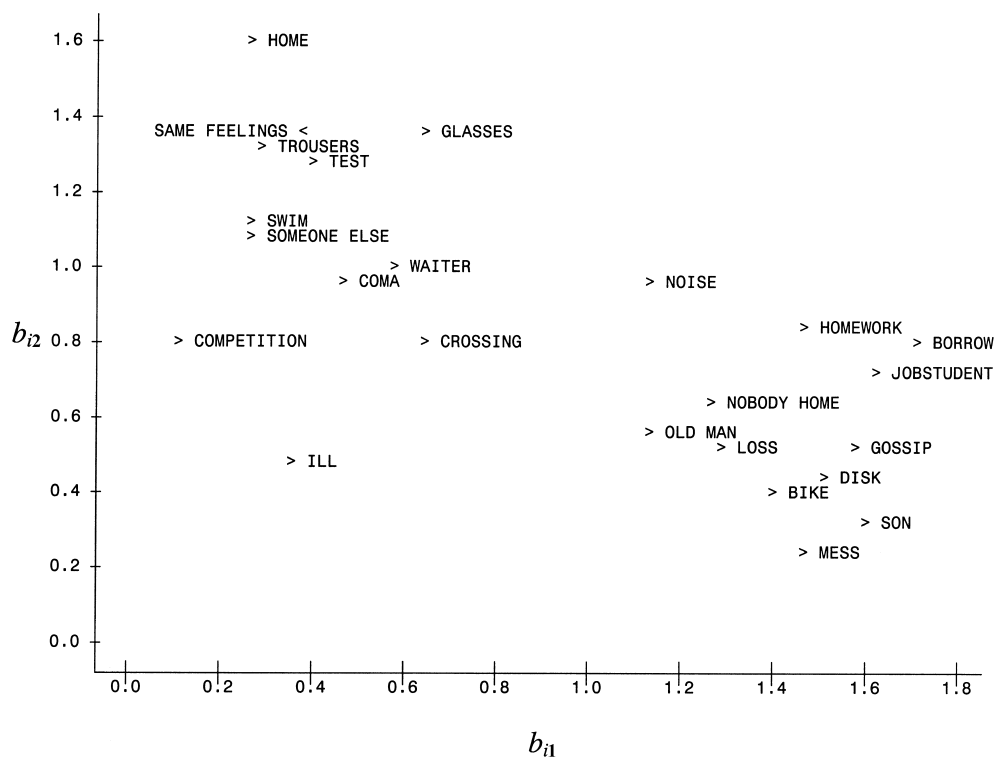


Figure 1. Loadings of the 24 situations on the two dimensions of the two-dimensional two-parameter logistic model.

mulated. However, instead of filling up the structure, we advocate that a psychometrician should define his or her own model, customized to his or her particular data set and/or the theory he or she is modeling. We followed the latter approach in the application on modeling anger, especially in the first analysis, in which the rating scale model was extended with a latent regression and a decomposition of the item parameters.

References

- Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1–23.
- Adams, R. J., Wilson, M., & Wu, M. L. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22*, 47–76.
- Agresti, A. (1989). Tutorial on modeling ordered categorical response data. *Psychological Bulletin, 89*, 290–301.
- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.
- Agresti, A., Booth, J. G., Hobert, J. P., & Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology, 30*, 27–80.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29–51.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM-algorithm. *Psychometrika, 46*, 443–459.
- Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor-analysis. *Applied Psychological Measurement, 12*, 261–280.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*, 153–168.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Society, 88*, 9–25.
- Breslow, N. E., & Lin, X. (1995). Bias correction in gen-

- eralised linear mixed models with a single component of dispersion. *Biometrika*, 82, 81–91.
- Butter, R., De Boeck, P., & Verhelst, N. (1998). An item response model with internal restrictions on item difficulty. *Psychometrika*, 63, 47–63.
- Clayton, D. (1996). Generalized linear mixed models. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo methods in practice* (pp. 275–303). New York: Chapman & Hall.
- Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman & Hall.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford, England: Oxford University Press.
- Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika*, 64, 407–433.
- Eysenck, H. J. (1953). Fragebogen als Messmittel der Persönlichkeit [Questionnaires as measures of personality]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1, 291–335.
- Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York: Springer-Verlag.
- Fischer, G. H. (1973). Linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359–374.
- Fox, J. P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 271–288.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman & Hall.
- Gilks, W. R., Thomas, A., & Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *The Statistician*, 43, 169–178.
- Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data. *Biometrika*, 78, 45–51.
- Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Edward Arnold.
- Goldstein, H., & Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*, 505–513.
- Hedeker, D. (1999). MIXNO: A computer program for mixed-effects nominal logistic regression [Computer software and manual]. Retrieved August 24, 2001, from <http://tigger.uic.edu/%7Ehedeker/manuals.html>
- Hedeker, D., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933–944.
- Hedeker, D., & Gibbons, R. D. (1996). MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, 49, 157–176.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.
- Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological Methods*, 2, 261–277.
- Jannarone, R. (1986). Conjunctive item response theory kernels. *Psychometrika*, 51, 357–373.
- Janssen, R., & De Boeck, P. (2003). *A random-effects version of the linear logistic test model*. Manuscript submitted for publication.
- Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38, 79–93.
- Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika*, 49, 223–245.
- Kuppens, P. (2003). *Individual differences in appraisal and emotion: The case of anger and irritation*. Unpublished manuscript.
- Lange, K., Hunter, D. R., & Yang, I. (2000). Optimization transfer using surrogate objective functions (with discussion). *Journal of Computational and Graphical Statistics*, 9, 1–52.
- Legler, J. M., & Ryan, L. M. (1997). Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association*, 92, 13–20.
- Lin, X., & Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, 91, 1007–1016.
- Longford, N. T. (1993). *Random coefficient models*. New York: Oxford University Press.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maier, K. (2001). The Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, 26, 307–330.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. London: Chapman & Hall.
- McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.

- McKinley, R. L. (1989). *Confirmatory analysis of test structure using multidimensional item response theory* (Research Report No. RR-89-31). Princeton, NJ: Educational Testing Service.
- McKinley, R. L., & Reckase, M. D. (1983). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. *Behavior Research Methods and Instrumentation*, *15*, 389–390.
- Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, *19*, 91–100.
- Mislevy, R. J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement*, *11*, 81–91.
- Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different persons employ different solution strategies. *Psychometrika*, *55*, 195–215.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176.
- Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*, 12–35.
- Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM manual* (Tech. Rep. No. 2001/01) [Software manual]. London: University of London, Department of Biostatistics and Computing.
- Rasbash, J., Browne, W. J., Goldstein, H., & Yang, M. (2000). *A user's guide to MLwiN (Version 2.1)* [Software manual]. London: Institute of Education.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis method* (2nd ed.). Newbury Park, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2001). *HLM 5* [Computer software and manual]. Lincolnwood, IL: Scientific Software.
- Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*, *9*, 141–157.
- Rijmen, F., & De Boeck, P. (2002). The random weights linear logistic test model. *Applied Psychological Measurement*, *26*, 269–283.
- Rodriguez, G., & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary response. *Journal of the Royal Statistical Society, Series A*, *158*, 73–89.
- Rosenbaum, P. R. (1988). Item bundles. *Psychometrika*, *53*, 349–359.
- Rosenberg, M. (1989). *Society and the adolescent self-image* (Rev. ed.). Middletown, CT: Wesleyan University Press.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271–282.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, *34*(Suppl., No. 17, pp. 100–114).
- Scott, S. L., & Ip, E. H. (2002). Empirical Bayes and item-clustering effects in a latent variable hierarchical model: A case study from the National Assessment of Educational Progress. *Journal of the American Statistical Association*, *97*, 409–419.
- Self, S. G., & Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*, 605–610.
- Skrondal, A., & Rabe-Hesketh, S. (in press). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*.
- Snijders, T., & Bosker, R. (1999). *An introduction to basic and advanced multilevel modeling*. London: Sage.
- Spiessens, B., Verbeke, G., Fieuws, S., & Rijmen, F. (2003). *Classification of clustered data using a SAS-macro: An application to latent class models*. Manuscript submitted for publication.
- Spiessens, B., Verbeke, G., & Komárek, A. (2002). *The use of mixed models for longitudinal count data when the random-effects distribution is misspecified*. Manuscript submitted for publication.
- StataCorp. (2001). *Stata statistical software* (Release 7) [Computer software and manual]. College Station, TX: Stata Press.
- Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed-effects model. *Biometrics*, *50*, 1171–1177.
- Tanner, M. A. (1996). *Tools for statistical inference* (3rd ed.). New York: Springer-Verlag.
- Thissen, D. (1991). *MULTILOG* [Software manual]. Lincolnwood, IL: Scientific Software.
- Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika*, *49*, 501–519.
- Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*, 567–577.
- Tuerlinckx, F., De Boeck, P., & Lens, W. (2002). Measuring needs with the Thematic Apperception Test: A psy-

- chometric study. *Journal of Personality and Social Psychology*, 82, 448–461.
- Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, 43, 39–55.
- Van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (1982). *Zelf-analyse vragenlijst* [Anger trait scale]. Lisse, The Netherlands: Swets & Zeitlinger.
- Verbeke, G., & Molenberghs, G. (1997). *Linear mixed models in practice: A SAS-oriented approach. Lecture notes in statistics 126*. New York: Springer-Verlag.
- Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag.
- Verguts, T., & De Boeck, P. (2000). A Rasch model for learning while solving an intelligence test. *Applied Psychological Measurement*, 24, 151–162.
- Verhelst, N. D., & Glas, C. A. W. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, 58, 395–415.
- Wolfinger, R. D., & Lin, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics and Data Analysis*, 25, 465–490.
- Wu, M. L., Adams, R. J., & Wilson, M. (1998). ACER ConQuest: Generalized Item Response Modelling Software [Computer software and manual]. Melbourne, Victoria, Australia: Australian Council for Educational Research.
- Yamamoto, K. (1987). *A hybrid model for item responses*. Unpublished doctoral dissertation, University of Illinois at Chicago.
- Zeger, S. L., & Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86, 79–86.
- Zwinderman, A. H. (1991). A generalized Rasch model for manifest predictors. *Psychometrika*, 56, 589–600.

Received January 3, 2002

Revision received September 9, 2002

Accepted December 19, 2002 ■

New Editor Appointed to *Psychological Review*, 2004–2010

The Publications and Communications Board of the American Psychological Association announces the appointment of Keith Rayner, PhD, to a 6½-year term on *Psychological Review* beginning on July 1, 2004.

Effective July 1, 2003, manuscript submissions to *Psychological Review* should be submitted electronically through the Manuscript Submission Portal at

<http://www.apa.org/journals/rev.html>

Authors who are unable to do so may contact the editor's office about alternatives by writing to Keith Rayner, PhD, Department of Psychology, Tobin Hall, University of Massachusetts, Amherst, MA 01003.

The current editor, Walter Mischel, PhD, will receive and consider manuscripts through June 30, 2003. Should the April 2004 issue be filled before that date, manuscripts will be redirected to the new editor for consideration.

Appendix A

Estimation of the Rating Scale Model With Latent Regression and Decomposition of the Item Parameters With the SAS NLMIXED Procedure

The model was estimated using the following code:

```

PROC NLMIXED data=ratscal method = gauss technique=newrap noad
qpoints=20;
PARMS intcpt = 0 beta_anger =0 beta_self= 0 beta_irritation = 0
beta_gender=0
      beta_control=0 beta_change=0 beta_predict=0 beta_conseq3=0
      beta_conseq1=0 beta_threat=0 beta_blame=0 beta_touch=0
      beta_clear=0 beta_loss=0 delta2 =0 delta3 =0 st2=1;

beta= intcpt+beta_anger*anger+beta_self*self+beta_irritation*irritation
      +beta_gender*gender+beta_control*control+beta_change*change
      +beta_predict*predict+beta_conseq3*conseq3+beta_conseq1*conseq1
      +beta_threat*threat+beta_blame*blame+beta_touch*touch+beta_clear*
clear
      +beta_loss*loss;

ex1=th1+beta;
ex2=2*(th1+beta)+delta2;
ex3=3*(th1+beta)+delta2+delta3;

if (y=0)then ex=0;
else if (y=1) then ex=ex1;
else if (y=2) then ex=ex2;
else ex=ex3;

p=exp(ex)/(1+exp(ex1)+exp(ex2)+exp(ex3));
ll=log(p);
MODEL y~general(ll);
RANDOM th1~normal(0 ,st2) subject = pp;
RUN;

```

The NLMIXED procedure is started with the PROC NLMIXED statement. The “data=” option specifies the data set. The first 30 lines of the SAS data set, named “ratscal,” are presented in Figure A1.

pp is an indicator variable for the participants. *y* is a column vector of length $510 \times 24 = 12240$ that results from stacking the 510 individual response vectors one below the other. *Anger*, *self*, *irritation*, and *gender* are person covariates, the other 10 variables are item covariates. Note that the number of rows of the dataset equals 12,240 and not $3 \times 12,240 = 36,720$, the number of rows of **X** (see text). This is because the different response categories of *y* are not recoded into three dichotomous variables but instead are treated separately in the program code. For the same reason, no category covariates are included either. This way, the size of the data set was kept much smaller.

The nonadaptive (“noad”) Gaussian quadrature method (“method = gauss”) was chosen to approximate the likelihood numerically, with 20 nodes (“qpoints=20”). The optimization algorithm was the Newton–Raphson technique (“technique=newrap”). Starting values for the parameters were specified in the PARMS statement. In the following lines, the loglikelihood (*ll*) of an individual response is defined and

parsed to the NLMIXED procedure by the MODEL statement. This procedure is followed because a multinomial distribution, which is the conditional distribution of y , is currently not available in the NLMIXED procedure. Finally, the RANDOM statement defines the random effect to be thl and specifies it to follow a normal distribution with mean zero and variance $st2$. “subject=pp” specifies that the distribution of the random effects is defined over the population of participants.

g e n d e r	0	b P s	P Y	i r r i t												
				c o r r e l a t i o n	p r o p o r t i o n	c o n t e n t	c o n t e n t	c o n t e n t	b e h a v i o r	t e c h n i c	c o m m u n i c a t i o n	a c a d e m i c	a c a d e m i c	S o c i a l	E d u c a t i o n	
1	1	1	1.84	1.44	0.920	2.40	0.24	0.88	0.36	0.48	1.52	0.48	1.0	2.3	2.1	
2	1	1	0.84	0.92	1.280	0.72	1.44	2.24	1.28	2.44	1.44	2.08	1.0	2.3	2.1	
3	1	1	1.24	1.96	1.160	0.88	0.68	1.16	0.76	1.16	1.92	1.32	1.0	2.3	2.1	
4	1	0	0.72	0.76	1.960	0.88	0.64	0.96	0.20	1.60	1.52	0.40	1.0	2.3	2.1	
5	1	3	0.92	1.24	1.600	1.84	2.20	2.24	0.28	2.24	2.00	1.12	1.0	2.3	2.1	
6	1	3	0.68	0.72	1.640	0.96	1.68	2.44	1.00	2.16	1.76	2.08	1.0	2.3	2.1	
7	1	2	0.56	1.36	0.800	0.64	1.36	1.76	1.88	1.60	2.00	1.00	1.0	2.3	2.1	
8	1	3	0.60	0.88	0.560	0.76	1.96	2.56	1.52	2.20	1.88	2.20	1.0	2.3	2.1	
9	1	2	1.16	1.44	1.920	1.72	1.00	1.68	0.36	1.36	1.84	0.96	1.0	2.3	2.1	
10	1	2	1.40	1.00	1.440	2.04	1.36	2.00	0.64	1.52	1.84	0.92	1.0	2.3	2.1	
11	1	2	0.80	0.80	2.240	1.12	1.20	1.56	1.92	2.20	1.48	1.36	1.0	2.3	2.1	
12	1	0	2.28	1.04	0.360	2.60	0.20	1.60	1.28	0.84	1.76	0.48	1.0	2.3	2.1	
13	1	3	1.12	1.60	2.200	1.88	0.40	0.56	0.72	1.24	1.84	0.32	1.0	2.3	2.1	
14	1	3	1.64	0.92	1.600	2.29	2.04	2.40	1.24	2.00	1.76	1.88	1.0	2.3	2.1	
15	1	2	0.88	0.56	0.360	1.32	1.76	2.20	1.60	2.16	1.32	1.88	1.0	2.3	2.1	
16	1	2	0.64	0.80	1.840	0.76	1.64	1.92	0.68	2.16	1.80	1.40	1.0	2.3	2.1	
17	1	2	0.68	0.80	0.840	1.76	1.72	2.08	0.48	1.32	1.56	1.16	1.0	2.3	2.1	
18	1	3	0.40	0.68	1.375	0.44	1.40	1.72	0.40	2.20	1.72	1.64	1.0	2.3	2.1	
19	1	2	1.04	0.92	0.760	1.04	1.40	1.72	1.40	1.44	1.60	1.20	1.0	2.3	2.1	
20	1	1	0.28	0.72	1.520	0.48	0.24	1.68	0.20	0.88	1.76	0.88	1.0	2.3	2.1	
21	1	3	0.44	0.64	1.480	0.64	1.36	1.76	0.32	2.08	1.28	1.64	1.0	2.3	2.1	
22	1	3	0.60	0.64	1.520	1.32	2.08	2.08	0.76	2.48	1.60	1.76	1.0	2.3	2.1	
23	1	3	0.12	0.16	2.080	0.28	1.16	1.76	0.24	2.44	1.28	2.16	1.0	2.3	2.1	
24	1	2	0.76	0.60	0.240	0.72	0.56	1.20	1.16	0.88	1.48	0.84	1.0	2.3	2.1	
25	2	1	1.84	1.44	0.920	2.40	0.24	0.88	0.36	0.48	1.52	0.48	1.1	1.8	1.5	
26	2	1	0.84	0.92	1.280	0.72	1.44	2.24	1.28	2.44	1.44	2.08	1.1	1.8	1.5	
27	2	1	1.24	1.96	1.160	0.88	0.68	1.16	0.76	1.16	1.92	1.32	1.1	1.8	1.5	
28	2	0	0.72	0.76	1.960	0.88	0.64	0.96	0.20	1.60	1.52	0.40	1.1	1.8	1.5	
29	2	2	0.92	1.24	1.600	1.84	2.20	2.24	0.28	2.24	2.00	1.12	1.1	1.8	1.5	
30	2	1	0.68	0.72	1.640	0.96	1.68	2.44	1.00	2.16	1.76	2.08	1.1	1.8	1.5	

Figure A1. The first 30 lines of the *ratscal* dataset for the rating scale model with latent regression and decomposition of the item location parameters.

Appendix B

Estimation of the Two-Dimensional Two-Parameter Logistic Model

The model was estimated using the following code:

```

PROC NL MIXED data=twopl method = gauss technique=quanew noad
qpoints=10;
PARMS beta_waiter=0 beta_someoneelse=0 beta_competition=0 beta_ill=0
      beta_son=0 beta_loss=0 beta_glasses=0
beta_disk=0 beta_mess=0
      beta_jobstudent=0 beta_crossing=0 beta_home=0 beta_oldman=0
      beta_borrow=0 beta_homework=0
      beta_samefeelings=0 beta_noise=0
      beta_nobodyhome=0 beta_test=0 beta_swim=0 beta_bike=0
beta_gossip=0      beta_coma=0 beta_trousers=0

      b_waiter1=1 b_someoneelse1=1 b_competition1=1 b_ill1=1 b_son1=1
      b_loss1=1 b_glasses1=1 b_disk1=1 b_mess1=1 b_jobstudent1=1
      b_crossing1=1 b_home1=1 b_oldman1=1 b_borrow1=1 b_homework1=1
      b_samefeelings1=1 b_noise1=1 b_nobodyhome1=1 b_test1=1 b_swim1=1
      b_bike1=1 b_gossip1=1 b_coma1=1 b_trousers1=1 b_someoneelse2=1
      b_competition2=1 b_ill2=1 b_son2=1 b_loss2=1 b_glasses2=1
      b_disk2=1 b_mess2=1 b_jobstudent2=1 b_crossing2=1 b_home2=1
      b_oldman2=1
      b_borrow2=1 b_homework2=1 b_samefeelings2=1 b_noise2=1
      b_nobodyhome2=1
      b_test2=1 b_swim2=1 b_bike2=1 b_gossip2=1 b_coma2=1
b_trousers2=1;

b1 = b_waiter1*waiter + b_someoneelse1*someoneelse +
      b_competition1*competition + b_ill1*ill + b_son1*son +
      b_loss1*loss
      + b_glasses1*glasses + b_disk1*disk + b_mess1*mess +
      b_jobstudent1*jobstudent + b_crossing1*crossing + b_home1*home +
      b_oldman1*oldman + b_borrow1*borrow + b_homework1*homework +
      b_samefeelings1*samefeelings + b_noise1*noise +
      b_nobodyhome1*nobodyhome + b_test1*test + b_swim1*swim +
      b_bike1*bike +      b_gossip1*gossip + b_coma1*coma +
      b_trousers1*trousers;

b2= 1*waiter+b_someoneelse2*someoneelse + b_competition2*competition
+
      b_ill2*ill + b_son2*son + b_loss2*loss +b_glasses2*glasses +
      b_disk2*disk + b_mess2*mess + b_jobstudent2*jobstudent +
      b_crossing2*crossing + b_home2*home + b_oldman2*oldman +
      b_borrow2*borrow + b_homework2*homework +
      b_samefeelings2*samefeelings + b_noise2*noise +
      b_nobodyhome2*nobodyhome + b_test2*test + b_swim2*swim +
      b_bike2*bike +      b_gossip2*gossip + b_coma2*coma +
      b_trousers2*trousers;

beta= beta_waiter*waiter + beta_someoneelse*someoneelse +
      beta_competition*competition + beta_ill*ill + beta_son*son +

```

```

        beta_loss*loss +beta_glasses*glasses + beta_disk*disk +
        beta_mess*mess + beta_jobstudent*jobstudent +
beta_crossing*crossing +
        beta_home*home + beta_oldman*oldman + beta_borrow*borrow +
        beta_homework*homework + beta_samefeelings*samefeelings +
        beta_noise*noise + beta_nobodyhome*nobodyhome + beta_test*test +
        beta_swim*swim + beta_bike*bike + beta_gossip*gossip +
beta_coma*coma +
        beta_trousers*trousers;

ex=exp(th1*b1+th2*b2+beta);
p=ex/(1+ex);
model y~binary(p);
random th1 th2 ~normal([0,0],[1,0,1]) subject=pp;
RUN;

```

The first 30 lines of the SAS data set, named “twopl,” are presented in Figure B1. *pp* is an indicator variable for the participants. *y* is a column vector of length $510 \times 24 = 12,240$ that results from stacking the 510 individual response vectors one below the other. The other variables are dummies coding for the items.

The nonadaptive (“noad”) Gaussian quadrature method (“method = gauss”) was chosen to approximate the likelihood numerically (“method = gauss noad”), with 10 nodes per dimension (“qpoints=10”). The optimization algorithm was a quasi-Newton technique (“technique=quanew”), which turned out to be somewhat faster than the Newton–Raphson technique. Starting values for the parameters were specified in the PARMs statement. In the following lines, the probability of success (*p*) of an individual response is defined, and then it is parsed to the NLMIXED procedure by the MODEL statement that the conditional distribution of *y* is the Bernoulli distribution with probability *p*. Finally, the RANDOM statement defines the random effects to be *th1* and *th2* and specifies them to follow a bivariate normal distribution with mean zero, variances of one, and zero covariance. “subject=pp” specifies that the distribution of the random effects is defined over the population of participants.

Continued on next page

Obs	P	Y	waiter	sooneer	competitor	ill	son	loss	glass	disk	mess	jobstudent	crossing	home	oldman	brother	homework	same	friend	noody	height	swim	bike	gossip	coach	teachers
1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
13	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
14	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
15	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
16	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
17	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
19	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
21	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
22	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
23	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
24	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
25	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure B1. The first 30 lines of the *twopl* dataset for the 2-dimensional 2-parameter logistic model.