The R Package CDM for Cognitive Diagnosis Models

Ann Cathrice George
BIFIE Salzburg

Alexander Robitzsch
IPN Kiel

Thomas Kiefer
BIFIE Salzburg

Jürgen Groß
University of Hildesheim

Ali Ünlü
TU München

Abstract

This paper introduces the R package **CDM** for cognitive diagnosis models (CDMs). The package implements parameter estimation procedures for two general CDM frameworks, the generalized-deterministic input noisy-and-gate (G-DINA) and the general diagnostic model (GDM). It contains additional functions for analyzing data under these frameworks, like tools for simulating and plotting data, or for evaluating global model and item fit. The paper describes the theoretical aspects of implemented CDM frameworks and it illustrates the usage of the package with empirical data of the common fraction subtraction test by Tatsuoka (1984).

Keywords: R, cognitive diagnosis models, general diagnostic model, structured latent class models, skill diagnosis.

1. Introduction

In recent years educational research was characterized by an increasing demand of complex information on students' achievement. This may be caused by a growing interest in explaining the results of international comparative studies like the trends in international science study (TIMSS; Mullis, Martin, Ruddock, O'Sullivan, Arora, and Erberer 2005), the progress in international reading study (PIRLS; Mullis, Martin, Kennedy, and Foy 2007) and the programme for international student assessment (PISA; OECD 2010). It may also be caused by a strong need to explain the social and ethnic disparities detected in these studies (e.g., Mullis *et al.* 2007).

At present, the preferred method to model students' responses in an achievement test are item response theory (IRT) models (e.g., van der Linden and Hambleton 1997). These models deliver a quantification of the items' difficulties and the respondents' abilities through real-valued parameters which are measured on one or more continuous factors. In order to obtain qualitative interpretations of the parameters, each of the continuous factors may be partitioned into discrete hierarchically ordered competence levels through cut points by means of a so-called standard setting procedure (Cizek, Bunch, and Konns 2004). A discrete classification of the students' abilities into these competence levels may then serve as a kind of feedback, for example, for evaluating if students fulfill educational standard norms (e.g., Martin, Mullis, and Kennedy 2007).

Another method allowing diagnostic conclusions about students' abilities is a cognitive diagnosis model (CDM; Rupp, Templin, and Henson 2010), which comprises two steps: In a first qualitative step of a CDM analysis, educational and didactic experts have to subdivide the tested ability into different basic abilities, which are termed skills. For the description of the skills and their connections to the tested ability the experts often rely on information given in so-called competence models. In these competence models the skills may be defined through components in the process of task solving, for example, finding a common denominator in fractional arithmetics (Tatsuoka 1984), or they represent coarsely defined abilities, for example, the six basic skills *systems*, *classification*, *observation*, *measurement*, *prediction* and *data* (Ackerman, Zhang, Henson, and Templin 2006; Rupp *et al.* 2010, p. 31f.). After defining the number and content of the skills, the experts expose in a so-called Q-matrix which skills are required to master each item (Tatsuoka 1983). Then, in a second quantitative step of the CDM analysis, the students are classified into dichotomous latent skill classes, which predict their presence or absence of the before defined skills. The main results are threefold: Firstly, the distribution of the skill classes allows for statements how many students in the test population possess certain combinations of skills. Secondly, the skill mastery probabilities include information about the percentage of students in the test population possessing the individual skills. Thirdly, for each individual student a skill class is deduced which is called the student's skill profile and which predicts the possession or nonpossession of the individual skills. Together all three issues provide a solid empirical base for targeted pedagogical interventions both on the level of the test population and on the individual student's level. In contrast to information obtained from the IRT framework, the CDM results may be interpreted directly in terms of competences the respondents possess. For further methodical differences between the CDM and the IRT models see for example Kunina-Habenicht, Rupp, and Wilhelm (2009).

CDMs have several connections to well-established areas in psychometrics: firstly, the theory of classification, where basic ideas can be found in the mastery model by Macready and Dayton (1977) and in restricted latent class models by Haertel (1989). Secondly, IRT with initial approaches in the multicomponent model by Whitely (1980) and in the linear logistic test model by Fischer (1973), and thirdly, in mathematical psychology, and here especially the field of knowledge space theory, see for example Doignon and Falmagne (1999). Based on the multitude of different approaches, CDMs have many names, as for example diagnostic classification models, cognitive psychometric models, or structured IRT models.

The CDM framework incorporates a huge variety of models which differ basically in three aspects: firstly, the combination in which students have to possess the skills for successfully mastering an item, that is, compensability: In some CDMs all assigned skills have to be possessed for mastering the items (noncompensatory models), in other CDMs just one

of the assigned skills has to be possessed, and other CDMs require one of several special combinations of the assigned skills (compensatory models). Secondly, CDMs vary in their item dimensionality: CDMs in which exactly one skill is required for mastering each item are called CDMs with between item dimensionality, whereas CDMs requiring more than one skill per item have a within item dimensionality. Finally, CDMs differ in the way the stochastic component is introduced into the model, that is, students can slip or guess in items, in skills, or in both. In an achievement test each item may follow another CDM. The different CDM rules for the items are sometimes also called the items condensation rules or simply the item rules (DiBello, Roussos, and Stout 2006).

Furthermore, CDM approaches may be partitioned into specific CDMs and generalized frameworks. The latter allow specifying many of the specific CDMs by constraining their parameters and defining their link functions. There are three common generalized CDM frameworks: the generalized-deterministic input noisy-and-gate (G-DINA; de la Torre 2011) model, the generalized diagnostic model (GDM; von Davier 2008) and the log-linear cognitive diagnosis model (LCDM; Henson, Templin, and Willse 2009).

The present work introduces and discusses the R (R Core Team 2016) package **CDM** (Robitzsch, Kiefer, George, and Ünlü 2016) for estimating and analyzing CDMs. The package contains routines for the estimation of the noncompensatory deterministic input noisy-and-gate (DINA; Haertel 1989; Junker and Sijtsma 2001) model, the compensatory deterministic input noisy-or-gate (DINO; Junker and Sijtsma 2001) model, and their generalized version, the G-DINA. These three model types and their parameter estimation process are reviewed in Sections 2.2, 2.3, and 2.4. Section 2.5 reviews the additional possibilities of the GDM and gives first prospects of its implementation. The LCDM framework is not exclusively implemented as it is very similar to one form of the G-DINA. Further implemented methods for the analysis of the before mentioned CDM frameworks are described in Section 2.6. In Section 3 the structure of the R package **CDM** is described, and Section 4 illustrates all models and methods with data of the fraction subtraction test by Tatsuoka (1984). Finally, Section 5 sheds light on parameter recovery and speed of the R package **CDM**.

2. Theory

2.1. Terminology and notation

Consider an achievement test in which I students respond to J items. A value of 1 indicates a correct item response and a value of 0 an incorrect one. The dichotomous manifest response of student i, $i = 1, \ldots, I$, to item j, $j = 1, \ldots, J$, is denoted by X_{ij}. The empirical responses of all I students to all J items are given in a $I \times J$ binary data matrix $\boldsymbol{X}$. The i-th row $\boldsymbol{X}_i$ of $\boldsymbol{X}$ represents the responses of student i to all J items and is called the i-th student's response pattern.

Educational experts define K ($K \leq J$) skills α_k, $k = 1, \ldots, K$, which students have to possess for mastering each of the J items under consideration. For each student i a latent dichotomous skill profile $\boldsymbol{\alpha}_i = [\alpha_{i1}, \ldots, \alpha_{iK}]$ denotes his possession ($\alpha_{ik} = 1$) and nonpossession ($\alpha_{ik} = 0$) of the K predefined skills. Obviously, these individual skill profiles are unknown (it is one goal of the CDM analysis to estimate them). Educational experts also define which skills are required to master which item in a $J \times K$ matrix $\boldsymbol{Q}$ (Tatsuoka 1983): The (j, k)-th element

q_{jk} of $\boldsymbol{Q}$ equals 1 if skill k is relevant for the mastery of item j and equals 0 otherwise. Additionally, the experts have to specify the items' condensation rules (i.e., the specific CDM each item follows).

CDMs assume that the manifest response X_{ij} of student i to item j arises as a result of his possessed skills i, the skills required for item j defined in the j-th row $\boldsymbol{q}_j$ of the Q-matrix, and the j-th item's condensation rule. Because the skills α_k are unknown, a CDM algorithm deduces from the manifest responses, the Q-matrix, and the condensation rules information on the K skills the student possesses.

This procedure comprises two steps: In a first step, all students are classified into skill classes $\boldsymbol{\alpha}_l$, $l = 1, \ldots, L$, satisfying a global optimization criterion. Note that the largest possible number of disjunctive skill classes is $L = 2^K$, arising from all combinations of the assumed K skills. From this first step the distribution of the skill class probabilities, that is, the relative frequencies $P(\boldsymbol{\alpha}_l)$, $l = 1, \ldots, L$, of students classified into the skill classes $\boldsymbol{\alpha}_l$, is obtained. Further, we obtain the skill mastery probabilities $P(\alpha_k)$, $k = 1, \ldots, K$, giving for each skill α_k the relative frequency of students in possession of it. Both the distribution of the skill classes and the skill mastery probabilities allow for diagnostic statements about the test population, that is, which skills are possessed in which combination and frequency.

In a second step, the CDM algorithm deduces the most probable skill classes each individual i, $i = 1, \ldots, I$, belongs to. The i-th student's vector of present and absent skills is also called the i-th student's *skill profile* and is denoted by $\boldsymbol{\alpha}_i = [\alpha_{i1}, \ldots, \alpha_{iK}]$. The skill profiles provide the basis for individual based diagnostic feedback, forming a solid empirical base for further instruction and learning.

Note that we have to distinguish between the skill classes $\boldsymbol{\alpha}_l$, $l = 1, \ldots, L$, in the population and the individual skill profiles $\boldsymbol{\alpha}_i$, $i = 1, \ldots, I$, both being K-length dichotomous vectors. The L possible skill classes cover all I skill profiles. For the ease of notation we use the same symbol "$\boldsymbol{\alpha}$". It will always become clear from the context whether $\boldsymbol{\alpha}_1$ refers to the first skill class or the first individual skill profile.

2.2. DINA and DINO

Because of its simplicity and its parsimony the DINA (Haertel 1989; Junker and Sijtsma 2001, function `din` in the R package) model is one of the most commonly used and most popular core CDMs. The DINA model's noncompensability asserts that students have to possess all skills assigned to an item for successfully mastering it. On the contrary, the compensatory DINO (Junker and Sijtsma 2001, function `din`) model requires students to possess at least one of the assigned skills for successfully mastering the respective item.

In both, DINA and DINO models, the i-th student's probability to master the j-th item involves two components: a deterministic and a probabilistic one. The former states whether the student is expected to master the j-th item given his possessed skills. In the DINA model the deterministic component is expressed through the dichotomous latent response

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$$

of student i with skill profile $\boldsymbol{\alpha}_i = [\alpha_{i1}, \ldots, \alpha_{iK}]$ to item j. The vector $[q_{j1}, \ldots, q_{jK}]$ again denotes the q-th row of the Q-matrix which indicates the skills required for the mastery of

item j. A student who possesses all or even more than these required skills is expected to master the item and $\eta_{ij} = 1$. Otherwise, for a student who is not expected to master the item $\eta_{ij} = 0$. In the DINO model, the deterministic latent response

$$\eta_{ij} = 1 - \prod_{k=1}^{K} (1 - \alpha_{ik})^{q_{jk}}$$

of student i to item j equals 1 if the student possesses at least one of the required skills. The second, that is the probabilistic component, possible deviates from these expectations: On the one hand, if student i is expected to master the item, he nevertheless may slip and fail the item. On the other hand, even if the student is not expected to master the item, he may succeed by a lucky guess. The probabilities g_j of guessing and s_j of slipping are modeled as item specific parameters with

$$g_j = \mathsf{P}(X_{ij} = 1 | \eta_{ij} = 0), \quad j = 1, \ldots, J$$

and

$$s_j = \mathsf{P}(X_{ij} = 0 | \eta_{ij} = 1), \quad j = 1, \ldots, J.$$

The probability of student i to solve item j results from combining both components

$$\mathsf{P}(X_{ij} = 1 | \boldsymbol{\alpha}_i, g_j, s_j) = (1 - s_j)^{\eta_{ij}} \cdot g_j^{(1 - \eta_{ij})} = \begin{cases} 1 - s_j & \text{for } \eta_{ij} = 1, \\ g_j & \text{for } \eta_{ij} = 0. \end{cases} \tag{1}$$

Hence, the probability $\mathsf{P}(X_{ij} = 1 | \boldsymbol{\alpha}_i, g_j, s_j)$ is modeled by only two values: All students who are not expected to master the item (i.e., $\eta_{ij} = 0$) have the chance g_j to solve the item by lucky guess, and all students who are expected to master the item (i.e., $\eta_{ij} = 1$) have the chance $1 - s_j$ to actually solve the item without slipping.

2.3. Parameter estimation in the DINA and DINO model

Parameter estimation of the DINA and DINO model is performed by means of marginal maximum likelihood (MML) estimation. A pertinent way to implement this method is the expectation-maximization (EM; Dempster, Laird, and Rubin 1977) algorithm (de la Torre 2009b). The EM algorithm iterates between an E-step and an M-step: In the E-step, expected counts for each item and each group are calculated. Then, the M-step updates the parameter estimates for the DINA or DINO model using maximization methods. Finally, the E-step and M-Step alternate until a previously set convergence criterion or a fixed maximum number of iterations is attained. Note that the process of parameter estimation is carried out in the same way if different items follow different condensation rules. Note also that for reasons of simplicity we present the parameter estimation process by the example of the DINA model (i.e., the DINO model, depending on the definition of the dichotomous latent response η_{ij}), but it may be extended to more complex CDMs in a straightforward manner (Xu and von Davier 2008a).

Up to this point, the probability $\mathsf{P}(X_{ij} = 1 | \boldsymbol{\alpha}_i)$ is interpreted as the probability of student i to master item j given his skills $\boldsymbol{\alpha}_i$. This notion facilitates the interpretation and the understanding of the models. But the students' skill profiles are unknown and the goal is to estimate them. Thus, we should understand $\mathsf{P}(X_{ij} = 1 | \boldsymbol{\alpha}_i)$ as the probability of student i to

master item j if he would be classified in skill class $l = i$, $l = 1, \ldots, L$. This convention is used throughout the following steps.

It is assumed that the responses $\boldsymbol{X}_i$ of student i to the items are independent conditional on $\boldsymbol{\alpha}_i$ (local independence assumption). Furthermore, it is assumed that examinees are mutually independent as well, because we expect them to represent a random sample of the population. Let $\boldsymbol{\delta} = [\boldsymbol{g}, \boldsymbol{s}]$ be the vector of all item parameters with $\boldsymbol{g} = [g_1, \ldots, g_J]$ and $\boldsymbol{s} = [s_1, \ldots, s_J]$. Furthermore let

$$\mathsf{P}\left(\boldsymbol{X}_i \middle| \boldsymbol{\alpha}_l, \boldsymbol{\delta}\right) = \prod_{j=1}^{J} \mathsf{P}\left(X_{ij} = 1 | \boldsymbol{\alpha}_l; g_j, s_j\right)^{X_{ij}} \left[1 - \mathsf{P}\left(X_{ij} = 1 | \boldsymbol{\alpha}_l; g_j, s_j\right)\right]^{1 - X_{ij}}$$

be the probability of response vector $\boldsymbol{X}_i$ if student i possesses the skills of skill class $\boldsymbol{\alpha}_l$, $l = 1, \ldots, L$.

For estimating the DINA model, the marginal log-likelihood

$$\log L(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \sum_{i=1}^{I} \log L(\boldsymbol{X}_i; \boldsymbol{\delta}, \boldsymbol{\gamma}) = \sum_{i=1}^{I} \log \left[\sum_{l=1}^{L} \mathsf{P}\left(\boldsymbol{X}_i | \boldsymbol{\alpha}_l; \boldsymbol{\delta}\right) \cdot \mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{\gamma})\right]$$

is maximized with respect to the item parameters $\boldsymbol{\delta}$ and the parameters $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_{L'}]$ describing the skill class distribution $\mathsf{P}(\boldsymbol{\alpha}_l), l = 1, \ldots, L$. In case of the estimation of an unrestricted skill space it holds $L = L'$ and $\gamma_l = \mathsf{P}(\boldsymbol{\alpha}_l)$ for all $l = 1, \ldots, L$. In case of an restricted skill space (i.e., through the application of a log-linear skill space or skill hierarchies) $L' \leq L - 1$. For the ease of presentation in the following we assume unrestricted skill class distributions.

Prior to the first iteration of the EM algorithm, initial item parameters $\boldsymbol{\delta}$ and skill distribution parameters $\boldsymbol{\gamma}$ have to be chosen. Then, the EM algorithm alternates between the E-step and the M-step described in the following:

E-Step

a. The individual posterior distribution can be deduced via Bayes' theorem:

$$\mathsf{P}(\boldsymbol{\alpha}_l \mid \boldsymbol{X}_i) = \frac{\mathsf{P}(\boldsymbol{X}_i \mid \boldsymbol{\alpha}_l)\,\mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{\gamma})}{\sum_{m=1}^{L} \mathsf{P}(\boldsymbol{X}_i \mid \boldsymbol{\alpha}_m)\,\mathsf{P}(\boldsymbol{\alpha}_m | \boldsymbol{\gamma})}, \quad l = 1, \ldots, L.$$

b. Two types of expected counts are derived from the posterior: The first count is the expected number

$$T_{jl} = \sum_{i=1}^{I} \mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{X}_i; \boldsymbol{\delta})$$

of students which are classified into skill class $\boldsymbol{\alpha}_l$ for item j, $j = 1, \ldots, J$. Note that in case of no missing data $T_{jl} = T_{j'l}$ for all j, $j' = 1, \ldots, J$. The second count

$$R_{jl} = \sum_{i=1}^{I} X_{ij} \cdot \mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{X}_i; \boldsymbol{\delta})$$

describes the expected number of students classified in skill class $\boldsymbol{\alpha}_l$ while responding item j correctly.

M-Step

a. The item parameter vector $\boldsymbol{\delta} = [\boldsymbol{g}, \boldsymbol{s}]$ is updated. The estimating equations are obtained by setting the first derivative of the expected complete log-likelihood with respect to the item parameters equal to zero. The derivative only involves the two counts obtained in the E-step. Let

$$T_j^{(0)} = \sum_{l=1}^{L} T_{jl} \cdot (1 - \eta_{lj})$$

be the expected number of students lacking at least one of the skills required for the mastery of item j (i.e., $\eta_{lj} = 0$) and

$$R_j^{(0)} = \sum_{l=1}^{L} X_{ij} \cdot T_{jl} \cdot (1 - \eta_{lj})$$

be the expected number of students among $T_j^{(0)}$ who correctly respond to item j. Furthermore, let $T_j^{(1)}$ and $R_j^{(1)}$ have the same interpretation except that they belong to students who possess all skills required for item j (i.e., $\eta_{lj} = 1$). Based on this definitions, the items parameters of item j are updated according to

$$\hat{g}_j = \frac{R_j^{(0)}}{T_j^{(0)}} \quad \text{and} \quad \hat{s}_j = 1 - \frac{R_j^{(1)}}{T_j^{(1)}}.$$

b. The skill class distribution $\mathsf{P}(\boldsymbol{\alpha}_l; \boldsymbol{\gamma})$ is updated. Therefore the expected number n_l of students in skill class $\boldsymbol{\alpha}_l$ is calculated, namely

$$n_l = \sum_{i=1}^{I} \mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{X}_i; \boldsymbol{\delta}).$$

Then the skill class distribution is updated by

$$\mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{\gamma}) = \frac{n_l}{I} \quad l = 1, \dots, L.$$

and the skill mastery probabilities are defined as

$$\mathsf{P}(\alpha_k) = \sum_{l=1}^{L} \alpha_{lk} \cdot \mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{\gamma}), \quad k = 1, \dots, K.$$

The E- and M-Step alternate until convergence. Convergence may be achieved if the maximal change between the parameter estimates or the relative change in the deviance is below a specific predefined value or after a maximum number of iterations. Finally, the covariance matrix for the estimated $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\gamma}}$ parameters (i.e., the empirical standard errors) is computed using the empirical cross-product approach (Paek and Cai 2014). As an other possibility, statistical inference based on resampling methods (jackknife, repeared replicate weights, bootstrap) can be conducted by applying `IRT.jackknife` to a fitted model.

Once the algorithm converged, the individual student classifications or *individual skill profiles* can be deduced from the probabilities $\mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{X}_i)$, according to three methods: Firstly, following

the approach of maximum a priori (MAP) classification, the largest value of $\mathsf{P}(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)$ gives the skill class into which student i is classified:

$$\hat{\boldsymbol{\alpha}}_{i;MAP} = \arg\max_{\boldsymbol{\alpha}_l} \{\mathsf{P}(\boldsymbol{\alpha}_l \mid \boldsymbol{X}_i)\}.$$

Secondly, an individual classification of student i based on maximum likelihood estimation (MLE) is obtained by maximizing

$$\hat{\boldsymbol{\alpha}}_{i;MLE} = \arg\max_{\boldsymbol{\alpha}_l} \{\mathsf{P}(\boldsymbol{X}_i \mid \boldsymbol{\alpha}_l)\}.$$

Finally, for a classification of student i based on expected a posteriori probabilities (EAP), the marginal skill probability $\mathsf{P}(\alpha_k \mid \boldsymbol{X}_i)$ of student i for mastering skill k is computed as the sum of all $\mathsf{P}(\boldsymbol{\alpha}_l \mid \boldsymbol{X}_i)$ corresponding to mastery of skill k (i.e., having a 1 in the k-th element)

$$\mathsf{P}(\alpha_k \mid \boldsymbol{X}_i) = \sum_{l=1}^{L} \alpha_{lk} \cdot \mathsf{P}(\boldsymbol{\alpha}_l \mid \boldsymbol{X}_i), \quad k = 1, \ldots, K.$$

Then, the i-th student's EAP skill profile is estimated by

$$\tilde{\boldsymbol{\alpha}}_{i;EAP} = [\,\mathsf{P}(\alpha_1 \mid \boldsymbol{X}_i), \ldots, \mathsf{P}(\alpha_K \mid \boldsymbol{X}_i)\,].$$

For deducing the dichotomous skill class $\hat{\boldsymbol{\alpha}}_{i;EAP}$ in which student i is classified according to EAP, each marginal skill mastery probability $\mathsf{P}(\alpha_k|\boldsymbol{X}_i)$, smaller than 0.5 is set to 0, whereas each one larger or equal to 0.5 is set to 1. For a comparison of MAP, MLE, and EAP classification methods see Huebner and Wang (2011).

It remains noting three aspects concerning the estimation algorithm: Firstly, the algorithm may also handle sampling weights, which is not presented here. Secondly, in particular situations not all $L = 2^K$ skill classes have to be estimated, for example assumptions of hierarchy between the K skills may be implemented in predefining specific skill class occurrence probabilities $\mathsf{P}(\boldsymbol{\alpha}_l)$ to be zero. Thirdly, in cases where models have almost as many parameters as observations, which, consequently, would lead to weakly or nonidentifiable skill classes, Xu and von Davier (2008b) proposed to change from the unreduced skill space $\mathsf{P}(\boldsymbol{\alpha}_l)$, $l = 1, \ldots, L$, to a log-linear smoothed form of the skill space.

2.4. G-DINA

For relaxing this restrictive two-probability constraint, de la Torre (2011) introduced the G-DINA (function `gdin` in the R package) framework. In this model, students exhibiting different sets of required skills have different probabilities of mastering item j. The G-DINA model employs the item response function

$$\mathsf{P}\left(X_{ij} = 1 \middle| \boldsymbol{\alpha}_{j;i}^*, \boldsymbol{\delta}_j\right) = \delta_{j;0} + \sum_{k=1}^{K_j^*} \delta_{j;k}\alpha_{j;ik}^* + \sum_{k=1}^{K_j^*-1}\sum_{k'=k+1}^{K_j^*} \delta_{j;kk'}\alpha_{j;ik}^*\alpha_{j;ik'}^* + \ldots + \delta_{j;12\ldots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{j;ik}^*.$$
$$(2)$$

Here $\boldsymbol{\alpha}_{j;i}^* = [\alpha_{j;i1}^*, \ldots, \alpha_{j;iK_j^*}^*]$ is the shortened skill profile of student i, which includes only the skills relevant for item j. Thus $K_j^* = \sum_{k=1}^{K} q_{jk}$ represents the number of skills which are necessary for the mastery of item j, that is, K_j^* is the sum of ones in the j-th row of the Q-matrix. In

the following, for notational convenience and without loss of generality, let the first K_j^* of all K skills be the ones required for item j. Then the former skill profiles $\boldsymbol{\alpha}_i$ decompose into different reduced skill profiles depending on item j, which necessitates the notation of an additional item index in each skill profile. Furthermore $\boldsymbol{\delta}_j = [\delta_{j;0}, \delta_{j;1}, \ldots, \delta_{K_j^*;1}, \delta_{j;12}, \ldots, \delta_{j;12\ldots K_j^*}]$ includes the item parameters with respect to item j, with $\delta_{j;0}$ being an intercept parameter, $\delta_{j;k}$ representing first-order effects, $\delta_{j;kk'}$ second-order effects and so on.

If only first-order effects $\delta_{j;k}$ are modeled (i.e., all other parameters are defined to be zero), the resulting models are called G-DINA 1way models. In the same manner, G-DINA models with first-order effects $\delta_{j;k}$ and second-order interaction effects $\delta_{j;kk'}$ are called G-DINA 2way models. Note that other versions of the G-DINA model use the logit or log link instead of the identity link (cf. Equation 2) for modeling the response probabilities. Note also that many core CDMs, as for example the DINA and DINO model, may be derived from the G-DINA framework by restricting parameters.

2.5. Possibilities of the GDM

The GDM (von Davier 2008, function `gdm` in the R package) framework includes nearly all common CDMs for both dichotomous and polytomous response data. Furthermore, with GDMs polytomous and even (quasi-)continuous skills can be established. Hence, the class of GDMs also includes a partial credit model for polytomous response data as well as uni- and multidimensional IRT models. Furthermore, in this framework Q-matrices with polytomous entries may be handled. The estimation of GDMs is based on MML methods and is implemented in the R package **CDM** based on an EM algorithm (Xu and von Davier 2008b).

2.6. Some methods for the analysis of CDMs

Model fit

The R package **CDM** supports the evaluation of model fit via the Akaike information criterion (AIC; Akaike 1973, function `AIC`) and the Bayesian information criterion (BIC; Schwarz 1978, function `BIC`) or an χ^2 overall goodness-of-fit measure which is based on the idea of local dependence between the items (Chen and Thissen 1997, function `IRT.modelfit`).

The AIC and BIC indices follow the well known formulas

$$\text{AIC} = -2 \log L(\boldsymbol{X}) + 2 \cdot p$$

and

$$\text{BIC} = -2 \log L(\boldsymbol{X}) + \log I \cdot p,$$

where $\log L(\boldsymbol{X})$ denotes the maximal log-likelihood of the model and p represents the total number of estimated parameters. For example for DINA or DINO models with K underlying skills it holds $p = 2 \cdot J + L - 1$.

The χ^2 measure evaluates, roughly spoken, the difference between the observed and the model-predicted response probabilities: While the predicted response probabilities are calculated assuming local independence between the items, the observed probabilities may deviate from this assumption. Consequently, if the analyzed CDM model approximates the (dependencies in the) data well, the χ^2 measure provides low values.

Let $N(X_j = n, X_{j'} = m)$ be the observed absolute number of students' responses fulfilling $X_j = n$ and $X_{j'} = m$, $n, m = 0, 1$. Let

$$\hat{N}(X_j = n, X_{j'} = m) = \sum_{i=1}^{I} \sum_{l=1}^{L} \mathsf{P}(X_{ij} = n | \boldsymbol{\alpha}_l) \mathsf{P}(X_{ij'} = m | \boldsymbol{\alpha}_l) \mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{X}_i).$$

be the predicted absolute number of students in the same response category, where $\mathsf{P}(\boldsymbol{\alpha}_l | \boldsymbol{X}_i)$ is the individual posterior from the fitted model and $\mathsf{P}(X_{ij} = n | \boldsymbol{\alpha}_l) = \mathsf{P}(X_{ij} = n | \boldsymbol{\alpha}_l; \boldsymbol{\delta})$ denotes the DINA model's item response function. Then the difference between the observed and the predicted probabilities for two arbitrary items j and j', j, $j' = 1, \ldots, J$, $j \neq j'$, is defined by

$$\chi_{jj'}^2 = \sum_{n=0}^{1} \sum_{m=0}^{1} \frac{\left[N(X_j = n, X_{j'} = m) - \hat{N}(X_j = n, X_{j'} = m) \right]^2}{\hat{N}(X_j = n, X_{j'} = m)}.$$

This fit statistic is chi square distributed with one degree of freedom. As a criterion of absolute fit, the maximal $\chi_{jj'}^2$ is considered (Chen, de la Torre, and Zhang 2013) and compared to the $1 - \alpha^*$ quantile $\chi_{1-\alpha^*, 1}^2$ of the chi squared distribution with one degree of freedom. Here, α^* is the Holm corrected (Holm 1979) significance level of the desired level α.

As other methods for evaluating the absolute model fit the **CDM** package contains effect size measures (function `modelfit.cor.din`) as introduced by Chen and Thissen (1997), Chen *et al.* (2013) DiBello *et al.* (2006), Maydeu-Olivares (2013) and McDonald and Mok (1995).

Item fit

The R package **CDM** provides as an item fit statistic the so called root mean square error of approximation (item-fit RMSEA; Kunina-Habenicht, Rupp, and Wilhelm 2009, function `itemfit.rmsea`), which indicates how good an item coincides with the model. The item-fit RMSEA of item j compares the model-predicted item response probabilities $\mathsf{P}(X_j = 1 | \boldsymbol{\alpha}_l)$ with the predicted absolute number of correct responses $\hat{N}(X_j = 1 | \boldsymbol{\alpha}_l)$ in each skill class $\boldsymbol{\alpha}_l$:

$$\text{RMSEA}_j = \sqrt{\sum_{l=1}^{L} p(\boldsymbol{\alpha}_l) \left[\mathsf{P}(X_j = 1 | \boldsymbol{\alpha}_l) - \frac{\hat{N}(X_j = 1 | \boldsymbol{\alpha}_l)}{\hat{N}(X_j | \boldsymbol{\alpha}_l)} \right]^2}.$$

Here $p(\boldsymbol{\alpha}_l)$ is the frequency of students classified in skill class $\boldsymbol{\alpha}_l$ and $\hat{N}(X_j | \boldsymbol{\alpha}_l)$ the predicted total number of responses (i.e., correct and incorrect ones) to item j given by students in skill class $\boldsymbol{\alpha}_l$. Kunina-Habenicht *et al.* (2009) recommend to evaluate item fit according to the following rules of thumb: Items with item fit indices below .05 indicate good fit, items with RMSEA values below .10 demonstrate moderate fit, and item-fit indices with $\text{RMSEA}_j > .10$ imply poor fit.

Item discrimination index

In DINA and DINO models the additional constraint $g_j < 1 - s_j$ ensures that the probability of mastering an item in possession of all required skills without slipping is higher than the probability of guessing an item while lacking at least one required skill. Because this constraint is not considered in the estimation process it may be checked with the item discrimination index

$IDI_j = 1 - s_j - g_j$ (Lee, de la Torre, and Park 2012, function `summary.din`), where negative IDI values indicate a violation of this constraint. The IDI may also be seen as a diagnostic index, reporting for each item how it discriminates between students possessing all skills (i.e., having a response probability of $1 - s_j$) and students lacking in at least one skill (i.e., guessing with probability g_j). Thus, IDIs close to 1 signal a good discrimination or "diagnosticity" of the item, whereas IDI values close to 0 detect items with a low discrimination.

Classification criteria

In some studies it is not only of interest to assess the global model fit but also to evaluate how well the response behavior of individual students is described by the model. In these cases classification criteria as the model's classification accuracy and consistency might be utilized. Classification accuracy is a measure of how well individual students are correctly classified in their true skill classes, whereas classification consistency is a measure for the consistence of the classifications in two parallel test forms with the same items and parameters. In the R package **CDM**, the classification accuracy and consistency (function `cdm.est.class.accuracy`) are assessed via simulation methods (cf. DiBello *et al.* 2006) and analytically by the method of Cui, Gierl, and Chang (2012). Concerning the former, the simulation is conducted with the estimated model parameters $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\gamma}}$. Note that both, the accuracy and consistency measures, rely on the assumption that the specified model is the correct one.

Tetrachoric correlation between skills

It may also be of interest to analyze correlations between the estimated skills. The calculation is based on the assumptions that latent continuous variables are underlying the latent dichotomous skill variables and that, under bivariate normality assumptions, the correlation between two skill variables equals the correlation between the two underlying continuous variables. With $\mathbb{1}$ denoting the indicator function, this so-called tetrachoric correlation between two skills α_k and $\alpha_{k'}$ (cf. Templin, Henson, Templin, and Roussos 2008) is inferred on the basis of a two times two frequency table with

$$\mathsf{P}(\alpha_k = n, \alpha_{k'} = m) = \sum_{l=1}^{L'} \mathbb{1}\{\alpha_{lk} = n\}\mathbb{1}\{\alpha_{lk'} = m\}\mathsf{P}(\boldsymbol{\alpha}_l|\boldsymbol{\gamma}).$$

3. Structure of the package

The R package **CDM** provides the three main functions `din`, `gdina`, and `gdm` which allow the estimation of DINA and DINO, G-DINA and GDM models, respectively. For objects returned by the model estimation functions `dina`, `gdin`, and `gdm`, S3 methods for generics like print and summary as well as some additional functions for further analysis of the respective CDMs are available. The package includes the simulated example data sets `sim.dina` and `sim.dino` for DINA and DINO models while simultaneously it allows to construct response data from DINA, DINO, and G-DINA models by using the simulation tools `sim.din` and `sim.gdina`. Table 1 yields a scheme of the package structure, whereas the basic output values of the main functions `din`, `gdin`, and `gdm` and their descriptions are given in Table 2. The functions `attribute.patt`, `skill.patt`, and `pattern` are not defined for the `gdm` function because they are not generalizable for the case of nondichotomous data and skills.

Command	Description	din	gdina	gdm
`anova`	Likelihood ratio test	✓	✓	✓
`coef`	Model parameters	✓	✓	✓
`logLik`	Log-likelihood	✓	✓	✓
`plot`	Graphical illustration	✓	✓	✓
`print`	Model output	✓	✓	✓
`summary`	Summary of model	✓	✓	✓
`IRT.compareModels`	Model comparison (≥ 2 models)	✓	✓	✓
`IRT.factor.scores`	Estimated individual skill profiles $\hat{\boldsymbol{\alpha}}_i$	✓	✓	✓
`IRT.likelihood`	Individual likelihood $\mathsf{P}(\boldsymbol{X}_i\|\boldsymbol{\alpha}_l)$	✓	✓	✓
`IRT.jackknife`	Generation of standard errors	✓	✓	✓
`IRT.modelfit`	χ^2 goodness-of-fit	✓	✓	✓
`IRT.posterior`	Individual posterior $\mathsf{P}(\boldsymbol{\alpha}_l\|\boldsymbol{X}_i)$	✓	✓	✓
`cdm.est.class.accuracy`	Class. accuracy/consistency	✓	✓	
`itemfit.rmsea`	RMSEA item fit	✓	✓	✓
`skill.cor`	Tetrachoric correlation of skills	✓	✓	✓
`sim.din`	Simulation of DINA/O data	✓		
`sim.gdin`	Simulation of G-DINA data		✓	

Table 1: S3 methods from R base package for CDM objects (top), S3 methods from **CDM** package for CDM objects (middle), and functions (bottom) provided by the **CDM** package.

Command	Description	din	gdina	gdm
`Npars`	Number of estimated parameters	✓	✓	✓
`posterior`	Posterior skill class distribution	✓	✓	✓
`attribute.patt`	Skill class distribution	✓	✓	
`skill.patt`	Skill mastery probabilities	✓	✓	
`pattern`	Individual skill profiles	✓	✓	

Table 2: Main elements contained in the objects returned by `din`, `gdina`, and `gdm`.

4. Examples

The following examples are conducted using data of the well-known fraction-subtraction data (Tatsuoka 1984), which comprises dichotomous responses of $I = 536$ junior high school students to $J = 20$ fraction subtraction items. The items, the definition of the $K = 8$ underlying skills, and their assignment to the items are given in Table 3. In the R package **CDM**, the response data is included in the object `fraction.subtraction.data` and the respective Q-matrix in the `fraction.subtraction.qmatrix` object.

DINA

The estimation of a DINA model, in which it is required that the students possess all assigned skills for mastering the items, is conducted with the command

```
R> colnames(fraction.subtraction.qmatrix) <- paste("Skill", 1:8, sep = "_")
```

No.	Item	Skills	No.	Item	Skills
1	$\frac{5}{3} - \frac{3}{4}$	$\alpha_4, \alpha_6, \alpha_7$	11	$4\frac{1}{3} - 2\frac{4}{3}$	$\alpha_2, \alpha_5, \alpha_7$
2	$\frac{3}{4} - \frac{3}{8}$	α_4, α_7	12	$1\frac{1}{8} - \frac{1}{8}$	α_7, α_8
3	$\frac{5}{6} - \frac{1}{9}$	α_4, α_7	13	$3\frac{3}{8} - 2\frac{5}{6}$	$\alpha_2, \alpha_4, \alpha_5, \alpha_7$
4	$3\frac{1}{2} - 2\frac{3}{2}$	$\alpha_2, \alpha_3, \alpha_5, \alpha_7$	14	$3\frac{4}{5} - 3\frac{2}{5}$	α_2, α_5
5	$4\frac{3}{5} - 3\frac{4}{10}$	$\alpha_2, \alpha_4, \alpha_7, \alpha_8$	15	$2 - \frac{1}{3}$	α_1, α_7
6	$\frac{6}{7} - \frac{4}{7}$	α_7	16	$4\frac{5}{7} - 1\frac{4}{7}$	α_2, α_7
7	$3 - 2\frac{1}{5}$	$\alpha_1, \alpha_2, \alpha_7$	17	$7\frac{3}{5} - \frac{4}{5}$	$\alpha_2, \alpha_5, \alpha_7$
8	$\frac{2}{3} - \frac{2}{3}$	α_7	18	$4\frac{1}{10} - 2\frac{8}{10}$	$\alpha_2, \alpha_5, \alpha_6, \alpha_7$
9	$3\frac{7}{8} - 2$	α_2	19	$4 - 1\frac{4}{3}$	$\alpha_1, \alpha_2, \alpha_3, \alpha_5, \alpha_7$
10	$4\frac{4}{12} - 2\frac{7}{12}$	$\alpha_2, \alpha_5, \alpha_7, \alpha_8$	20	$4\frac{1}{3} - 1\frac{5}{3}$	$\alpha_2, \alpha_3, \alpha_5, \alpha_7$

Table 3: Assignment of 8 skills to 20 items of fraction subtraction test by Tatsuoka (1984). In the table the skills are denoted as follows: α_1 describes the skill of converting a whole number to a fraction, α_2 is separating a whole number from a fraction, α_3 denotes simplifying before subtracting, α_4 stands for finding a common denominator, α_5 indicates the skill of borrowing from whole part, α_6 expresses column borrowing to subtract the second numerator from the first, α_7 represents subtracting numerators, and α_8 stands for reducing answers to simplest form.

```
R> fractions.dina <- din(data = fraction.subtraction.data,
+    q.matrix = fraction.subtraction.qmatrix, rule = "DINA",
+    progress = TRUE)
```

With the command

```
R> summary(fractions.dina)
```

important information about the model is obtained, as for example the values and the number of the estimated model parameters, goodness-of-fit measures and the tetrachoric correlations between the skills. All model parameters can be accessed through

```
R> coef(fractions.dina)
```

Additionaly a comprehensive item parameter summary, including item parameters and the item-fit RMSEA measure, is stored in the object

```
R> fractions.dina$item
```

The following output is only given for the first 5 out of the altogether 20 items:

```
      type guess se.guess  slip se.slip  rmsea
Item1 DINA 0.030    0.002 0.089   0.016 0.0527
Item2 DINA 0.016    0.001 0.041   0.008 0.0562
Item3 DINA 0.000    0.000 0.134   0.024 0.0985
Item4 DINA 0.224    0.015 0.110   0.021 0.1489
Item5 DINA 0.301    0.036 0.172   0.019 0.1356
```

Furthermore, the tetrachoric correlations between the skills

```
R> skill.cor(fractions.dina)$cor.skills
```

	Skill_1	Skill_2	Skill_3	Skill_4	Skill_5	Skill_6	Skill_7	Skill_8
Skill_1	1.000	0.351	0.436	0.690	0.643	0.517	0.140	0.416
Skill_2	0.351	1.000	0.424	0.317	0.190	0.498	0.918	0.693
Skill_3	0.436	0.424	1.000	0.365	0.602	0.383	0.401	0.395
Skill_4	0.690	0.317	0.365	1.000	0.554	0.607	0.341	0.319
Skill_5	0.643	0.190	0.602	0.554	1.000	0.514	0.180	0.327
Skill_6	0.517	0.498	0.383	0.607	0.514	1.000	0.569	0.427
Skill_7	0.140	0.918	0.401	0.341	0.180	0.569	1.000	0.634
Skill_8	0.416	0.693	0.395	0.319	0.327	0.427	0.634	1.000

show that especially Skills 2 and 7 provide a very high correlation and may thus be not distinguishable given the 20 fraction subtraction items. Gathered from the skill class distribution

```
R> fractions.dina$attribute.patt
```

most students (36%) are classified into the skill class "11111111", in which they are predicted to master all skills. The skill mastery probabilities are given with

```
R> fractions.dina$skill.patt
```

	skill.prob
Skill_1	0.580
Skill_2	0.786
Skill_3	0.717
Skill_4	0.688
Skill_5	0.602
Skill_6	0.792
Skill_7	0.814
Skill_8	0.817

According to this result only 58% of the junior high school students are able to convert a whole number to a fraction (α_1), but 81% of the students master subtract numerators (α_7) and to reduce answers to simplest form (α_8). The method IRT.factor.scores extracts the individual student classifications, with for example,

```
R> IRT.factor.scores(fractions.dina, type = "MLE")[1, ]
```

```
MLE.skill1 MLE.skill2 MLE.skill3 MLE.skill4 MLE.skill5 MLE.skill6
         1          1          1          0          1          1
MLE.skill7 MLE.skill8
         1          1
```

specifying the MLE classification of the first student. Accordingly, the MAP and EAP classifications can be obtained by specifying the respective `type`. Thus, student 1 is classified into class "11101111" via MLE meaning that he probably lacks skill α_4. The classifications via MAP and EAP correspond. The main results of a DINA model, including the EAP classification of an arbitrary student (here: Student 1), may be illustrated by

```
R> par(mfrow = c(2,2))
R> plot(fractions.dina, pattern = "000100110111101110111", ask = FALSE)
```

All plots are shown in Figure 1. The first plot on the top left hand side of Figure 1 depicts the DINA model's item parameters, that is, the guessing and nonslipping probabilities for each item. The plot signalizes low discriminations for Items 5 and 8. The second plot on the top right hand side illustrates the skill mastery probabilities: Students most probably master skills α_8 *reduce answers*, α_7 *substract numerators* and α_6 *column borrow*. The third plot on the bottom left hand side shows the skill class distribution. Skill classes in which students are most frequently classified are labeled. The fourth plot analyzes the possessed skills of an individual student, in this case Student 1. According to the model Student 1 possesses all skills except of α_4 *find a common denominator*.

G-DINA

The command

```
R> fractions.gdina1 <- gdina(fraction.subtraction.data,
+   fraction.subtraction.qmatrix, rule = "GDINA1")
```

fits a G-DINA 1way model for the fraction subtraction data and the respective Q-matrix. The distribution of the skill classes, the skill mastery probabilities, and the individual skill profile for Student 1 may again be accessed via

```
R> fractions.gdina1$attribute.patt
R> fractions.gdina1$skill.patt
R> IRT.factor.scores(fractions.gdina1, type = "MLE")[1, ]
```

The full G-DINA model including all interaction effects is obtained with the command

```
R> fractions.gdina <- gdina(fraction.subtraction.data,
+   fraction.subtraction.qmatrix)
```

For selecting the more appropriate model, one may compare the goodness of fit by AIC, BIC, and χ^2 indices of the two models

```
R> AIC(fractions.gdina1)
R> BIC(fractions.gdina1)
R> IRT.modelfit(fractions.gdina1)
```

or conduct a likelihood ratio test:
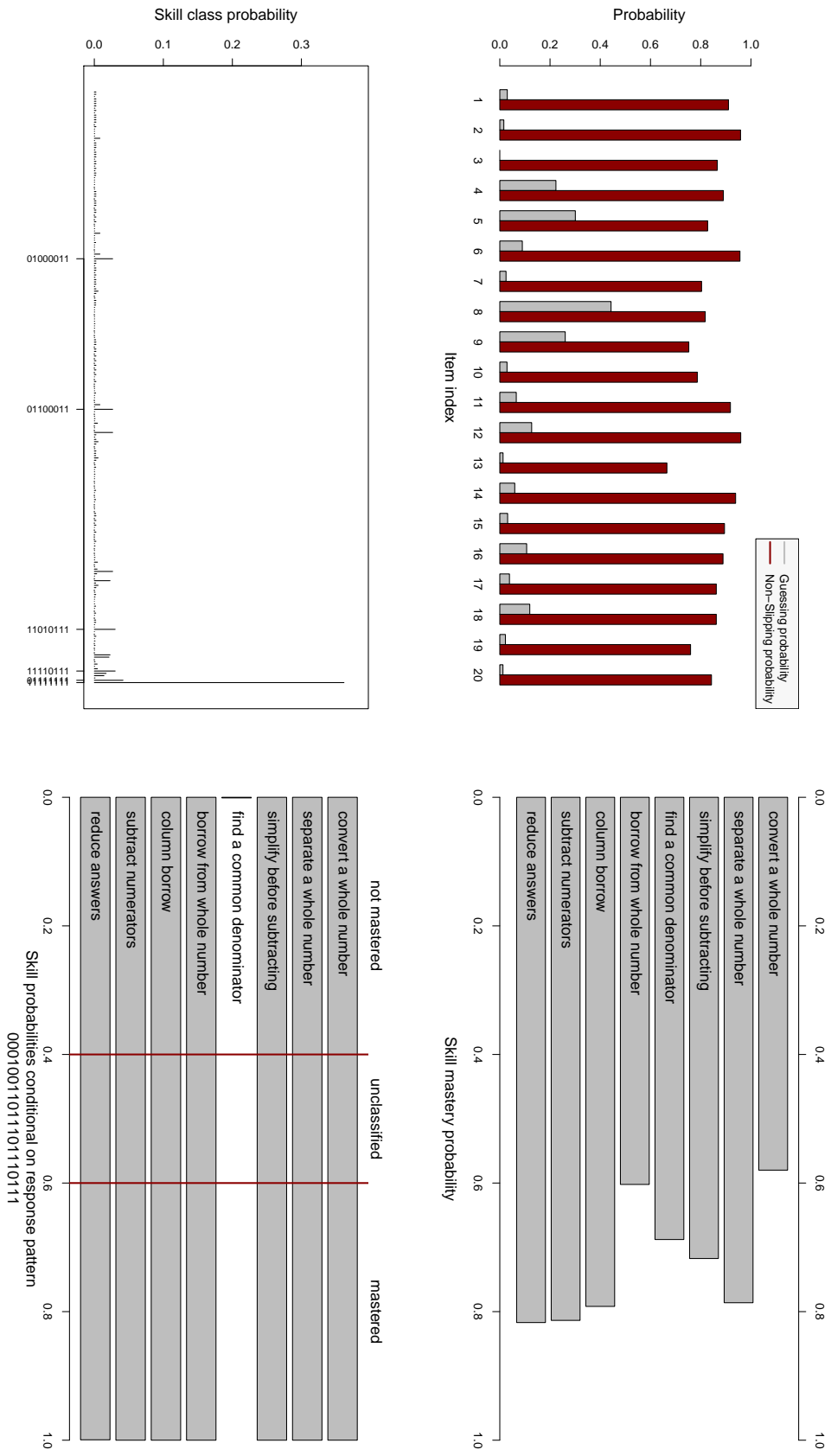
```
R> anova(fractions.gdina1, fractions.gdina)
```

Figure 1: Plot of results from DINA model `fractions.dina` with method `plot(fractions.dina)`: Item parameters (top, left hand side), skill mastery probabilities (top, right hand side), skill class distribution (bottom, left hand side) and individual EAP classification for Student 1 (bottom, right hand side).

GDM

As an example for fitting item response models in **CDM**, we estimate the Rasch model (Rasch 1960) by using the commands

```
R> theta.k <- seq(-6, 6, len = 15)
R> fractions.rasch <- gdm(fraction.subtraction.data, irtmodel = "1PL",
+    theta.k = theta.k, centerintercepts = TRUE)
```

where `theta.k` defines the cutpoints for the discretization of the continuous ability parameter. The model's summary shows, among others, the common item difficulty parameters `b.Cat1`.

The following output is again only given for the first 5 out of the altogether 20 items.

```
R> fractions.rasch$item[1:5,]
```

```
        item   N     M  b.Cat1 a.F1 itemfit.rmsea
Item1 Item1 536 0.532 -0.0285    1        0.0820
Item2 Item2 536 0.576  0.3193    1        0.0995
Item3 Item3 536 0.515 -0.1587    1        0.1044
Item4 Item4 536 0.528 -0.0575    1        0.1281
Item5 Item5 536 0.584  0.3776    1        0.1422
```

5. Parameter recovery and speed

This section focuses on parameter recovery issues of the EM algorithm for estimating CDMs which is implemented in the R package **CDM**. For illustrational purposes, DINA models including $J = 30$ items, $K = 5$ skills, and sample sizes of $I = 100, 300$, and 1000 students are simulated, with all guessing and slipping parameters varying between 0 and 0.29. The Q-matrix used for the simulation can be found in Table 1 of de la Torre (2009b). 5000 data sets are simulated for each of the three sample sizes. Each of the simulated data sets is fitted with a DINA model by setting the convergence criterion `conv.crit` of the `din` function to 0.0001. Table 4 gives the mean bias and the minimal and maximal absolute bias between the true and the estimated guessing, slipping, and skill class parameters. Table 4 also gives the mean, minimal, and maximal root mean square error (RMSE) between the true and the estimated guessing, slipping, and skill class parameters. The results indicate that the algorithm can provide accurate parameter estimates. All parameter estimates are unbiased and the RMSE decreases with increasing sample size. Parameter recovery for the above mentioned data sets is also tested with algorithms in other noncommercial software packages as for example the Ox (Doornik 2009) script by de la Torre (2009b), an lem (Vermunt 1997) procedure, and the mdltm software by von Davier (2008). The estimated item parameters and skill class probabilities of all the three before mentioned software packages and the R package **CDM** coincided up to the third or fourth position after the decimal point, which shows a reasonable agreement between the results. We may note that at the time of writing, no R function allowed the estimation of restricted latent class models as required for fitting CDMs.

Furthermore, we compared the speed of the **CDM** R algorithm with the Ox code for the DINA model provided by de la Torre (2009b). In both programs a DINA model is fitted to a

| | | | Bias | | | RMSE | |
Parameter type	I	mean	|min|	|max|	mean	min	max
Guessing g_j	100	−0.001	0.000	0.011	0.042	0.013	0.083
	300	0.000	0.000	0.003	0.024	0.007	0.046
	1000	0.000	0.000	0.001	0.013	0.004	0.025
Slipping s_j	100	0.000	0.000	0.005	0.059	0.005	0.100
	300	0.000	0.000	0.003	0.033	0.002	0.057
	1000	0.000	0.000	0.001	0.018	0.000	0.031
Skill class distribution $P(\boldsymbol{\alpha}_l)$	100	0.000	0.000	0.004	0.017	0.005	0.038
	300	0.000	0.000	0.001	0.009	0.003	0.022
	1000	0.000	0.000	0.001	0.005	0.018	0.012

Table 4: Mean bias, minimal and maximal absolute bias and mean, minimal and maximal RMSEA between the true and the estimated guessing, slipping and skill class parameters in a simulation study for parameter recovery in DINA models with $J = 30$ items, $K = 5$ skills and sample sizes of $I = 100, 300, 1000$.

simulated dataset with a sample size of $I = 1000$ persons, $J = 30$ items, and $K = 5$ attributes (for details of the simulated data see de la Torre 2009b) while the convergence criterion for item parameter change is set to 10^{-5}. On a desktop computer with a 2.67 GHz processor and 4 GB of memory, the **CDM** `din` function for R had a higher computational speed than the Ox script (R: 0.37 s, Ox: 1.04 s).

6. Conclusion

This paper introduces the R package **CDM**, covering parameter estimation procedures for DINA, DINO, G-DINA, and GDM models (`din`, `gdina`, `gdm`). Additionally, the package provides functions for analyzing data under these models: a simulation tool for generating data (`sim.din`, `sim.gdina`), functions for evaluating the global model fit or item fit (`IRT.modelfit`, `itemfit.rmsea`), and a method for describing the model's classification accuracy (`cdm.est.class.accuracy`). The methods contained in the package **CDM** have been described and their usage has been demonstrated by the empirical fraction subtraction data `fraction.subtraction.data` of Tatsuoka (1984). The package **CDM** further includes other data published data sets (e.g., `data.ecpe`; Templin and Hoffman 2013).

The R package **CDM** provides some more advanced methods which have not been discussed in this article: For example, it allows the estimation of multiple group designs (cf. Xu and von Davier 2008b) or the reduction of the skill space (cf. Xu and von Davier 2008a). In addition, the user may specify hierarchies on skills (`skillspace.hierarchy`; cf. Leighton, Gierl, and Hunka 2004; Templin and Bradshaw 2014), conduct a deterministic classification method (`din.deterministic`; Chiu, Douglas, and Li 2009), or empirically validate the Q-matrix (`din.validate.qmatrix`; de la Torre 2008). One may also evaluate CDMs concerning the test items' quality by means of the item discrimination index (`cdi.kli`; Henson, Roussos, Douglas, and He 2008), or concerning their uniqueness in terms of equivalent models (`equivalent.dina` von Davier 2014) or ambiguity of skill classes (`equivalent.dina` Groß and George 2014). Moreover, Wald tests can be employed for the item-specific choices of DINA or a G-DINA

rules (`gdina.wald`; de la Torre and Lee 2013) or for assessing differential item functioning (`gdina.dif`; Hou, de la Torre, and Nandakumar 2014).

Alternatively to conducting polytomous GDMs, it is further possible to estimate cognitive diagnosis models for polytomous attributes (`pgdina`; Chen and de la Torre 2013). Polytomous items can be handled by using the `gdm` function or by transforming the polytomous items in a set of dichotomous pseudo items (`sequential.items`; Tutz 1997), which are then deployed subsequently in DINA or GDINA models. Furthermore, multiple choice CDMs (de la Torre 2009a) may also be analyzed with the R package **CDM** as well as the generalized distance discrimination method (Sun, Xin, Zhang, and de la Torre 2013).

In principle, the implementation of the structured latent class model (`slca` Formann 1992; Formann and Kohlmann 1998) in the R package **CDM** covers most, if not all, of the implemented cognitive diagnostic models in the package. This function can also, and more generally, be used for the specification of constrained latent class analysis or many multidimensional item response models. For example, recent developments in the field of cognitive diagnostic models (e.g., Hong, Wang, Lim, and Douglas 2015; Huo and de la Torre 2014) may be implemented as a particular structured latent class model using the `slca` function (c.f. von Davier 2009).

References

Ackerman T, Zhang W, Henson R, Templin J (2006). "Evaluating a Third Grade Science Benchmark Test Using a Skills Assessment Model: Q-Matrix Evaluation." Paper presented at the Annual meeting of the National Council on Measurement in Education, San Fransisco.

Akaike H (1973). "Information Theory and an Extension of the Maximum Likelihood Principle." In B Petrov, F Csake (eds.), *Second International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado, Budapest.

Chen J, de la Torre J (2013). "A General Cognitive Diagnosis Model for Expert-Defined Polytomous Attributes." *Applied Psychological Measurement*, **37**, 419–437. `doi:10.1177/0146621613479818`.

Chen J, de la Torre J, Zhang Z (2013). "Relative and Absolute Fit Evaluation in Cognitive Diagnosis Modeling." *Journal of Educational Measurement*, **50**, 123–140. `doi:10.1111/j.1745-3984.2012.00185.x`.

Chen WH, Thissen D (1997). "Local Dependence Indexes for Item Pairs Using Item Response Theory." *Journal of Educational and Behavioral Statistics*, **22**, 265–289. `doi:10.3102/10769986022003265`.

Chiu CY, Douglas JA, Li X (2009). "Cluster Analysis for Cognitive Diagnosis: Theory and Applications." *Psychometrika*, **74**, 633–665. `doi:10.1007/s11336-009-9125-0`.

Cizek GJ, Bunch MB, Konns H (2004). "Setting Performance Standards: Contemporary Methods." *Educational Measurement: Issues and Practice*, **23**(4), 31–50. `doi:10.1111/j.1745-3992.2004.tb00166.x`.

Cui Y, Gierl MJ, Chang HH (2012). "Estimating Classification Consistency and Accuracy for Cognitive Diagnostic Assessment." *Journal of Educational Measurement*, **49**, 19–38. `doi:10.1111/j.1745-3984.2011.00158.x`.

de la Torre J (2008). "An Empirically Based Method of Q-Matrix Validation for the DINA Model: Development and Applications." *Journal of Educational Measurement*, **45**, 343–362. `doi:10.1111/j.1745-3984.2008.00069.x`.

de la Torre J (2009a). "A Cognitive Diagnosis Model for Cognitively Based Multiple-Choice Options." *Applied Psychological Measurement*, **33**, 163–183. `doi:10.1177/0146621608320523`.

de la Torre J (2009b). "DINA Model and Parameter Estimation: A Didactic." *Journal of Educational and Behavioral Statistics*, **34**, 115–130. `doi:10.3102/1076998607309474`.

de la Torre J (2011). "The Generalized DINA Model Framework." *Psychometrika*, **76**, 179–199. `doi:10.1007/s11336-011-9207-7`.

de la Torre J, Lee YS (2013). "Evaluating the Wald Test for Item-Level Comparison of Saturated and Reduced Models in Cognitive Diagnosis." *Journal of Educational Measurement*, **50**, 355–373. `doi:10.1111/jedm.12022`.

Dempster AP, Laird NM, Rubin DB (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society B*, **39**, 1–38.

DiBello LV, Roussos LA, Stout W (2006). "Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models." In CR Rao, S Sinharay (eds.), *Handbook of Statistics, Volume 26, Psychometrics*, pp. 979–1030. Elsevier, Amsterdam.

Doignon JP, Falmagne JC (1999). *Knowledge Spaces.* Springer-Verlag, Berlin. `doi:10.1007/978-3-642-58625-5`.

Doornik JA (2009). *Object-Oriented Matrix Programming Using Ox.* Timberlake Consultants Press, London. URL `http://www.doornik.com/`.

Fischer GH (1973). "The Linear Logistic Test Model as an Instrument in Educational Research." *Acta Psychologica*, **37**, 359–374. `doi:10.1016/0001-6918(73)90003-6`.

Formann AK (1992). "Linear Logistic Latent Class Analysis for Polytomous Data." *Journal of the American Statistical Association*, **87**, 476–486. `doi:10.2307/2290280`.

Formann AK, Kohlmann T (1998). "Structural Latent Class Models." *Sociological Methods & Research*, **26**, 530–565. `doi:10.1177/0049124198026004005`.

Groß J, George AC (2014). "On Permissible Attribute Classes in Noncompensatory Cognitive Diagnosis Models." *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, **10**, 100–107. `doi:10.1027/1614-2241/a000079`.

Haertel EH (1989). "Using Restricted Latent Class Models to Map the Skill Structure of Achievement Items." *Journal of Educational Measurement*, **26**, 301–321. `doi:10.1111/j.1745-3984.1989.tb00336.x`.

Henson RA, Roussos L, Douglas J, He X (2008). "Cognitive Diagnostic Attribute-Level Discrimination Indices." *Applied Psychological Measurement*, **32**, 275–288. doi:10.1177/0146621607302478.

Henson RA, Templin JL, Willse JT (2009). "Defining a Family of Cognitive Diagnosis Models Using Log-Linear Models with Latent Variables." *Psychmetrika*, **74**, 191–210. doi:10.1007/s11336-008-9089-5.

Holm S (1979). "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics*, **6**, 65–70.

Hong H, Wang C, Lim YC, Douglas J (2015). "Efficient Models for Cognitive Diagnosis with Continuous and Mixed-Type Latent Variables." *Applied Psychological Measurement*, **39**, 31–43. doi:10.1177/0146621614524981.

Hou L, de la Torre J, Nandakumar R (2014). "Differential Item Functioning Assessment in Cognitive Diagnostic Modeling: Application of the Wald Test to Investigate DIF in the DINA Model." *Journal of Educational Measurement*, **51**, 98–125. doi:10.1111/jedm.12036.

Huebner A, Wang C (2011). "A Note on Comparing Examinee Classification Methods for Cognitive Diagnosis Models." *Educational and Psychological Measurement*, **71**, 407–419. doi:10.1177/0013164410388832.

Huo Y, de la Torre J (2014). "Estimating a Cognitive Diagnostic Model for Multiple Strategies via the EM Algorithm." *Applied Psychological Measurement*, **38**, 464–485. doi:10.1177/0146621614533986.

Junker BW, Sijtsma K (2001). "Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory." *Applied Psychological Measurement*, **25**, 258–272. doi:10.1177/01466210122032064.

Kunina-Habenicht O, Rupp AA, Wilhelm O (2009). "A Practical Illustration of Multidimensional Diagnostic Skills Profiling: Comparing Results From Confirmatory Factor Analysis and Diagnostic Classification Models." *Studies in Educational Evaluation*, **35**, 64–70. doi:10.1016/j.stueduc.2009.10.003.

Lee YS, de la Torre J, Park YS (2012). "Relationships Between Cognitive Diagnosis, CTT, and IRT Indices: An Empirical Investigation." *Asia Pacific Education Review*, **13**, 333–345. doi:10.1007/s12564-011-9196-3.

Leighton JP, Gierl MJ, Hunka SM (2004). "The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule-Space Approach." *Journal of Educational Measurement*, **41**, 205–237. doi:10.1111/j.1745-3984.2004.tb01163.x.

Macready GB, Dayton CM (1977). "The Use of Probabilistic Models in the Assessment of Mastery." *Journal of Educational Statistics*, **2**, 99–120. doi:10.3102/10769986002002099.

Martin MO, Mullis IVS, Kennedy AM (eds.) (2007). *PIRLS 2006 Technical Report.* Boston College, Chestnut Hill.

Maydeu-Olivares A (2013). "Goodness-of-Fit Assessment of Item Response Theory Models." *Measurement: Interdisciplinary Research & Perspectives*, **11**, 71–101. doi:10.1080/15366367.2013.831680.

McDonald RP, Mok MMC (1995). "Goodness of Fit in Item Response Models." *Multivariate Behavioral Research*, **30**, 23–40. doi:10.1207/s15327906mbr3001_2.

Mullis IVS, Martin MO, Kennedy AM, Foy P (2007). *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries.* Boston College, Chestnut Hill.

Mullis IVS, Martin MO, Ruddock GJ, O'Sullivan CY, Arora A, Erberer E (2005). *TIMSS 2007 Assessment Frameworks.* Boston College, Chestnut Hill.

OECD (2010). *PISA 2009 Results: What Students Know and Can Do.* Organisation for Economic Co-operation and Development (OECD), Paris, France. doi:10.1787/9789264091450-en.

Paek I, Cai L (2014). "A Comparison of Item Parameter Standard Error Estimation Procedures for Unidimensional and Multidimensional Item Response Theory Modeling." *Educational and Psychological Measurement*, **74**, 58–76. doi:10.1177/0013164413500277.

Rasch G (1960). *Probabilitstic Models for Some Intelligence and Attainment Tests.* Ph.D. thesis, Danish Institute for Educational Research, Copenhagen, Denmark.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robitzsch A, Kiefer T, George AC, Ünlü A (2016). **CDM**: *Cognitive Diagnosis Modeling.* R package version 5.2-1, URL https://CRAN.R-project.org/package=CDM.

Rupp AA, Templin J, Henson RA (2010). *Diagnostic Measurement: Theory, Methods, and Applications.* The Guilford Press, New York.

Schwarz G (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, **6**, 461–464. doi:10.1214/aos/1176344136.

Sun J, Xin T, Zhang S, de la Torre J (2013). "A Polytomous Extension of the Generalized Distance Discriminating Method." *Applied Psychological Measurement*, **37**, 503–521. doi:10.1177/0146621613487254.

Tatsuoka KK (1983). "Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory." *Journal of Educational Measurement*, **20**, 345–354. doi:10.1111/j.1745-3984.1983.tb00212.x.

Tatsuoka KK (1984). "Analysis of Errors in Fraction Addition and Subtraction Problems." *Final Report for Grant No. NIE-G-81-0002*, University of Illinois, Urbana-Champaign, IL.

Templin J, Bradshaw L (2014). "Hierarchical Diagnostic Classification Models: A Family of Models for Estimating and Testing Attribute Hierarchies." *Psychometrika*, **79**, 317–339. doi:10.1007/s11336-013-9362-0.

Templin J, Hoffman L (2013). "Obtaining Diagnostic Classification Model Estimates Using Mplus." *Educational Measurement: Issues and Practice*, **32**(2), 37–50. `doi:10.1111/emip.12010`.

Templin JL, Henson RA, Templin SE, Roussos L (2008). "Robustness of Hierarchical Modeling of Skill Association in Cognitive Diagnosis Models." *Applied Psychological Measurement*, **32**. `doi:10.1177/0146621607300286`.

Tutz G (1997). "Sequential Models for Ordered Responses." In W van der Linden, RK Hambleton (eds.), *Handbook of Modern Item Response Theory*, pp. 139–152. Springer-Verlag, New York.

van der Linden WJ, Hambleton RK (eds.) (1997). *Handbook of Modern Item Response Theory*. Springer-Verlag, New York. `doi:10.1007/978-1-4757-2691-6`.

Vermunt JK (1997). *LEM: A General Program for the Analysis of Categorical Data*. Tilburg University, Netherlands.

von Davier M (2008). "A General Diagnostic Model Applied to Language Testing Data." *British Journal of Mathematical and Statistical Psychology*, **61**, 287–307. `doi:10.1348/000711007x193957`.

von Davier M (2009). "Some Note on the Reinvention of Latent Structure Models as Diagnostic Classification Models." *Measurement: Interdisciplinary Research & Perspective*, **7**, 67–74. `doi:10.1080/15366360902799851`.

von Davier M (2014). "The DINA Model as a Constrained General Diagnostic Model: Two Variants of a Model Equivalency." *British Journal of Mathematical and Statistical Psychology*, **67**, 49–71. `doi:10.1111/bmsp.12003`.

Whitely SE (1980). "Multicomponent Latent Trait Models for Ability Tests." *Psychometrika*, **45**, 479–494. `doi:10.1007/bf02293610`.

Xu X, von Davier M (2008a). "Comparing Multiple-Group Multinomial Log-Linear Models for Multidimensional Skill Distributions in the General Diagnostic Model (RR-08-35)." *Technical report*, Educational Testing Service, Princeton.

Xu X, von Davier M (2008b). "Fitting the Structured General Diagnostic Model to NAEP Data (RR-08-27)." *Technical report*, Educational Testing Service, Princeton.

Affiliation:

Ann Cathrice George
BIFIE – Federal Institute for Educational Research, Innovation &
Development of the Austrian School System
Alpenstraße 121
5020 Salzburg, Austria
E-mail: `a.george@bifie.at`
URL: `https://www.bifie.at/user/george-ann-cathrice`

Alexander Robitzsch
IPN – Leibniz Institute for Science and Mathematics Education at Kiel University
Olshausenstraße 62
24118 Kiel, Germany
E-mail: robitzsch@ipn.uni-kiel.de
URL: https://sites.google.com/site/alexanderrobitzsch