



The DINA model as a constrained general diagnostic model: Two variants of a model equivalency

Matthias von Davier*

Research & Development, Educational Testing Service, Princeton, New Jersey, USA

The 'deterministic-input noisy-AND' (DINA) model is one of the more frequently applied diagnostic classification models for binary observed responses and binary latent variables. The purpose of this paper is to show that the model is equivalent to a special case of a more general compensatory family of diagnostic models. Two equivalencies are presented. Both project the original DINA skill space and design \mathbf{Q} -matrix using mappings into a transformed skill space as well as a transformed \mathbf{Q} -matrix space. Both variants of the equivalency produce a compensatory model that is mathematically equivalent to the (conjunctive) DINA model. This equivalency holds for all DINA models with any type of \mathbf{Q} -matrix, not only for trivial (simple-structure) cases. The two versions of the equivalency presented in this paper are not implied by the recently suggested log-linear cognitive diagnosis model or the generalized DINA approach. The equivalencies presented here exist independent of these recently derived models since they solely require a linear – compensatory – general diagnostic model without any skill interaction terms. Whenever it can be shown that one model can be viewed as a special case of another more general one, conclusions derived from any particular model-based estimates are drawn into question. It is widely known that multidimensional models can often be specified in multiple ways while the model-based probabilities of observed variables stay the same. This paper goes beyond this type of equivalency by showing that a conjunctive diagnostic classification model can be expressed as a constrained special case of a general compensatory diagnostic modelling framework.

1. Introduction

The purpose of this paper is to show that the 'deterministic-input noisy-AND' (DINA) model (Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977) is equivalent to a special case of a more general compensatory family of diagnostic models.

Two variants of this model equivalency are presented in this paper. An empirical illustration of the result is given using example data that comes with the Ox (Doornik,

*Correspondence should be addressed to Dr. Matthias von Davier, Research & Development, Educational Testing Service, MS 13E, Princeton, New Jersey 08541, USA (email: mvondavier@ets.org). This paper was peer reviewed and accepted under the editorship of Thom Baguley.

2002) implementation of the DINA model (de la Torre, 2009). The equivalencies are based on defining two mappings of the DINA skill space and the design **Q**-matrix of item by skills associations. This (re)mapping produces an instance of the compensatory general diagnostic model (GDM; von Davier, 2005) that is mathematically equivalent to the (conjunctive) DINA model.

Whenever it can be shown that several equivalent model variants exist or one model can be viewed as a special case of another more general one, conclusions derived from any particular model-based estimates are called into question.

It is a widely known that multidimensional models can often be specified in multiple ways while the model-based probabilities of observed variables stay the same. Maris and Bechger (2004, 2009) have shown for MIRID ('model with internal restrictions on the item difficulties'; Butter *et al.*, 1998) models and diagnostic models that there are often multiple design (**Q**-matrix) and skill-space definitions for the same diagnostic model, and that these produce the same model-based probabilities of observed quantities. For the testlet model (Bradlow *et al.*, 1999) and the (constrained) bifactor model (Gibbons & Hedeker, 1992), this was shown, for example, by Rijmen (2010). For structural equation models (SEMs), it is a widely known result that the covariance matrix can be reproduced by rather different SEMs in identical or almost identical fashion. Finally, the equivalency between higher-order factor models and the hierarchical factor model has been established by Yung *et al.* (1999).

This paper goes beyond this type of equivalency by showing that a *conjunctive* diagnostic classification model can be expressed as a special case embedded in a general *compensatory* diagnostic modelling framework. More specifically, this paper looks at a different type of equivalency between cognitive diagnosis models. While previous research was mainly concerned with transformational or rotational invariance, this paper shows that a conjunctive diagnostic classification model can be expressed as a constrained model in a compensatory modelling framework – the GDM (von Davier, 2005, 2008, 2010).

The equivalencies presented in this paper hold for all DINA models with any **Q**-matrix, not only for trivial (simple-structure) cases. Also, these equivalencies are in no way implied by the recently suggested log-linear cognitive diagnosis model (L-CDM; Henson *et al.*, 2009) – derived on the basis of the GDM (Rupp *et al.*, 2010) – or the generalized DINA (G-DINA; de la Torre, 2011) approaches. The skill interaction features that these models add to the linear GDM are neither used nor approximated in the results presented here. These equivalencies solely require a linear – compensatory – GDM without any skill interaction terms. That is, in the result presented here, no additional structures are introduced, and both the DINA and the GDM are taken in their original form. An equivalency between the two is established purely by remapping the skill space of the DINA into an alternative skill space of a DINA-equivalent GDM.

At some level it may seem trivial to show that the vast majority of models considered 'diagnostic' can be considered latent class, or better, latent structure models (von Davier, 2009; von Davier *et al.*, 2008; von Davier & Yamamoto, 2004; Rupp *et al.*, 2010). But it is less obvious that a *conjunctive* model (i.e., a model that does not allow for compensatory functioning of skills) can be re-expressed using a mapping of the DINA skill space onto a new set of skills so that a *compensatory* diagnostic model can be used to define a DINA-equivalent model.

The long-standing knowledge that all diagnostic models can be viewed as latent structure models or generalized latent variable models and that therefore, at a very high level, these models are 'all the same' is not of much help to gain a deeper understanding of them. At this level of generality, it is also true for the mixture item response theory (IRT) and multidimensional IRT models, but it is still necessary to develop equivalency results

with regard to these models in order to facilitate understanding of the similarities and differences between these approaches (Rijmen & De Boeck, 2005).

Another example is the seminal paper by Takane and de Leeuw (1987): while factor analysis for discrete variables and IRT are both examples of a general class of latent variable models, it is important to develop a deeper understanding of the relationship between the specific instances of this general class.

The point of the results presented here is related to these types of published research on model equivalencies: within this class of latent structure models, different model assumptions can be made that lead to very different interpretations of diagnostic models, as in the case of assumed conjunctive or compensatory skills. If an equivalency is proven between these two apparently incompatible assumptions, this is new knowledge not explained by the fact that at a high level, both are instances of general latent variable models. To make an analogy, it is new knowledge if a researcher finds out that a rodent and a bear are genetically closer than their differences in appearance suggest even though it has been known for a long time that both are mammals.

As an added value gained en passant, we show how the equivalent-DINA model contains parameter constraints as well as constraints on the distribution of skills, and these results facilitate decisions as to whether the DINA model or some other, more general, diagnostic model is appropriate to fit the data at hand.

The selection of a particular model should be based on an examination of how well model assumptions relate to the theoretical consideration that served as the basis for test construction. By selecting and committing to use the DINA model early on, without considering whether other models may be more appropriate, the researcher skips this important step in determining whether one or more models are suitable representation(s) of the construct of interest. With this early decision comes an inability to examine whether the restrictions used in the DINA model are suitable or whether a more general model should have been used.

Embedding the DINA model – or, better, embedding the equivalent-DINA model variants into a larger modelling framework – allows comparisons to other models. The famous adage that all models are wrong while some may be useful (Box & Draper, 1987) also holds true for diagnostic models. Therefore, any (diagnostic) model is at best an approximation of reality, and a more general basis upon which different models can be compared is useful in determining whether the assumed skill requirements for each item, or even more so, whether the assumed functioning of skills as compensatory or non-compensatory/conjunctive, are indeed appropriate. It is expected that the result presented in this paper facilitates these types of model comparisons.

2. Some diagnostic classification models

For an overview of diagnostic classification models, the reader is referred to Fu and Li (2007), von Davier *et al.* (2008), and Rupp *et al.* (2010). Instead of giving a full overview of recent developments, this section describes only those models that are relevant for the equivalency result presented in this paper.

2.1. The DINA model

The DINA model is said to be conjunctive because it reduces the respondent-skill by item-attribute comparison to only two levels. Only those respondents who possess all necessary skills have a ‘high’ probability of solving an item, while respondents who lack at

least one of the required skills have a ‘low’ probability – the same ‘low’ probability no matter how many or which skills are missing.

The DINA model can be introduced more formally in the following way. Let I, N, K be integers denoting the number of items $i = 1, \dots, I$, the number of respondents $v = 1, \dots, N$, and the dimension of a latent variable $\mathbf{a} = (a_1, \dots, a_k)$, respectively. For each item i and each respondent v there is a binary (observable) response variable $X_{vi} \in \{0, 1\}$ where 0 represents a correct response and 1 an incorrect response. We will refer to $\mathbf{a} = (a_1, \dots, a_k)$ as the skill pattern in the following. For diagnostic models, we often assume that the components of this vector-valued latent variable are binary, $a_k \in \{0, 1\}$, indicating the absence or presence of skills $k = 1, \dots, k$. Note, however, that polytomous ordered skill variables can be used as well (von Davier, 2005). For each item, let

$$\mathbf{q}_i = (q_{i1}, \dots, q_{ik}),$$

with $q_{ik} \in \{0, 1\}$, define the vector of required skills; that is, $q_{ik} = 1$ if skill k is required for item i and $q_{ik} = 0$ otherwise. Then define the ‘conjunction function’ for respondent v and item i as

$$\eta_{vi} = f(\mathbf{q}_i, \mathbf{a}_v) = \prod_{k=1}^k a_{vk}^{q_{ik}}.$$

This function is based on the skill vector of the respondent $\mathbf{a}_v = (a_{v1}, \dots, a_{vk})$ and the vector of required skills $\mathbf{q}_i = (q_{i1}, \dots, q_{ik})$ and takes values $\eta_{vi} = 1$ if respondent v has all required skills for item i , and $\eta_{vi} = 0$ otherwise.

Finally, if the DINA model holds, the probability of a correct response for respondent v and item i can be written as

$$P(X_{vi} = 1 | \eta_{vi}, g_i, s_i) = g_i^{1-\eta_{vi}} (1 - s_i)^{\eta_{vi}},$$

where g_i is the *guessing* probability for item i quantifying the rate at which a person who does not possess all required skills will produce a correct response on item i . The parameter s_i denotes the *slipping* probability, quantifying the rate at which a respondent who possesses all required skills nevertheless produces an incorrect response on item i .

Note that the g_i and s_i denote item parameters, so that there are two parameters per item in the DINA model. In addition, the skill vectors $\mathbf{a}_v = (a_{v1}, \dots, a_{vk})$ are unobserved, so we typically have to assume the distribution of skills $P(\mathbf{A} = (a_1, \dots, a_k)) = \pi_{(a_1, \dots, a_k)}$ is unknown. Therefore, there are $\|\{0, 1\}^k\| - 1 = 2^k - 1$ independent skill pattern probabilities with $\sum_{(a_1, \dots, a_k)} \pi_{(a_1, \dots, a_k)} = 1.0$ if an unconstrained estimate of the skill distribution is attempted. There may be fewer parameters if a constrained distribution over the skill space (von Davier & Yamamoto, 2004; Xu & von Davier, 2008a,b) is used. For model identification, no constraints are needed on the guessing and slipping parameters (even though it is desirable that both are less than 0.5 for sensible results).

While de la Torre (2009) does not make statements about identifiability of the DINA model and the uniqueness of the model parameters, Junker and Sijtsma (2001) talk only

about (a lack of) empirical identification in the context of their data example used in conjunction with Markov chain Monte Carlo (MCMC) estimation. Haertel (1989) talks about identification of latent class skill patterns in the DINA model, and notes that ‘it may be impossible to distinguish all these classes empirically using a given set of items. Depending upon the items’ skill requirements, latent response patterns for two or more classes may be identical’ (p. 303). One of the remedies Haertel (1989) suggests is the combination of two or more latent classes that cannot be distinguished.

Because the DINA as described by these authors is estimated without skill distribution constraints, we will reduce the complexity of the estimation problem in the analysis section from 32 latent skill patterns used by de la Torre (2009) to be estimated separately to 16 distinct skill patterns. The approach taken will be outlined in more detail in the empirical section of this paper, together with a discussion of how to assess identifiability using the model equivalencies developed here. The decision to reduce the number of skills, and with that the number of skill patterns, should lead to a better-conditioned estimation problem and reduce identification issues as pointed out by Haertel (1989).

2.2. The general diagnostic model

The GDM (von Davier, 2005, 2008; von Davier & Yamamoto, 2004) provides a framework for the development of diagnostic models. As an item response modelling framework, the probability of an item response $x \in \{0, \dots, m_i\}$ by respondents $v = 1, \dots, N$ on items $i = 1, \dots, I$ can be written as

$$P(X = x|i, v) = \frac{\exp(f(\lambda_{xi}, \theta_v))}{1 + \sum_{y=1}^{m_i} \exp(f(\lambda_{xi}, \theta_v))}, \quad (1)$$

with item parameters $\lambda_{xi} = (\beta_{xi}, \mathbf{q}_i, \gamma_{xi})$ and a skill vector $\theta_v = (a_{v1}, \dots, a_{vk})$ with either continuous, ordinal, or, as in the case of the DINA and most other diagnostic models, binary skill variables $a_{.k} \in \{0, 1\}$. While this general form of the model served as the basis for numerous developments, among others, the L-CDM (e.g., Rupp *et al.*, 2010) for binary skill attributes and data, von Davier (2005, 2008) used the general form to derive the linear or partial-credit GDM

$$P(X = x|i, v) = \frac{\exp(\beta_{ix} + \sum_{k=1}^k \gamma_{ixk} b(q_{ik}, a_k))}{1 + \sum_{y=1}^{m_i} \exp(\beta_{ix} + \sum_{k=1}^k \gamma_{ixk} b(q_{ik}, a_k))} \quad (2)$$

with discrete skill vector $\mathbf{a}_{.k} \in \{a_{k1}, a_{k2}, \dots, a_{kL_k}\}$ which may contain ordinal or binary components, and $b(q, a) = qa$ and $\gamma_{ixk}^* = x\gamma_{ik}$ for parsimony. Note that these choices lead to a model that contains located latent class, multiple classification latent class, IRT, and multidimensional IRT models, as well as a compensatory version of the reparameterized unified model as special cases (von Davier, 2005, 2008; von Davier *et al.*, 2011). In addition, the linear GDM as well as the family of GDMs are suitable for binary, polytomous ordinal, and mixed-format item response data.

The model defined in equation (2) uses a weighted linear combination of skill components and is therefore a compensatory model by design, while the general framework (von Davier & Yamamoto, 2004) given in equation (1) can be used to define

compensatory as well as non-compensatory and conjunctive models. The result presented here, however, uses the linear GDM and shows how a conjunctive model is equivalent to a compensatory, linear GDM.

The GDM can have as many as $K + m_i$ parameters per item, up to K slopes and m_i thresholds. In addition, the skill space needs to be modelled. This is typically done by means of applying a log-linear model to predict the skill space probabilities (von Davier & Yamamoto, 2004) that is found to improve stability of estimation and greatly reduce the threat of an unidentified estimate of the skill space distribution (Xu & von Davier, 2008a). However, in the way the GDM is used in this paper, for binary items and skills and, more specifically, to define two different DINA-equivalent models, the equivalent model variants will have the same number of parameters per item as the DINA. The identification conditions mentioned above therefore apply in the same way.

3. Reparameterization A of the DINA model

In this section, a reparameterization is developed that allows the DINA to be recast as a special case of the linear (compensatory) GDM. This approach produces a DINA-equivalent GDM that provides the same model-based probabilities as the DINA. The reparameterization is based on defining a new set of skills, skill patterns, and a non-simple-structure **Q**-matrix for the GDM. This section shows that the reparameterization is equivalent to the DINA model.

For $t = 1, \dots, 2^k$, let a_t denote the t th skill pattern in the original K -dimensional DINA skill space. Then for any t , let

$$a_{ts}^* = \begin{cases} 1, & \text{for } s = \sum_{k=1}^k 2^{k-1} a_{tk}, \\ 0, & \text{otherwise.} \end{cases}$$

This defines a set of 2^k skill patterns which we will denote by $T(\mathbf{a})$. Table 1 shows an example with $k = 3$ skills and $8 = 2^3$ skill vectors of the DINA and the associated eight equivalent skill vectors for the reparameterized model.

For this first DINA-equivalent model, the definition of the transformed **Q**-matrix is carried out as follows. Let $\mathbf{q}_i = (q_{i1}, \dots, q_{ik}) \in \{0, 1\}^k \setminus \{(0, \dots, 0)\}$ denote the skill requirement vector for item i for the DINA model. Then define

Table 1. Skill attribute patterns for a DINA model with three dimensions and the equivalent skill attribute patterns for the compensatory GDM

DINA \mathbf{a}				Mapped into skill pattern \mathbf{a}^*						
	A	B	C	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0	0	0	0
2	+	0	0	+	0	0	0	0	0	0
3	0	+	0	0	+	0	0	0	0	0
4	+	+	0	0	0	+	0	0	0	0
5	0	0	+	0	0	0	+	0	0	0
6	+	0	+	0	0	0	0	+	0	0
7	0	+	+	0	0	0	0	0	+	0
8	+	+	+	0	0	0	0	0	0	+

Table 2. Three-dimensional \mathbf{Q} -matrix from a DINA model and the equivalent seven-skill \mathbf{Q} -matrix rows for the compensatory GDM

	DINA q			Remapped equivalent q^*						
	A	B	C	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0	0	0	0
2	+	0	0	+	0	+	0	+	0	+
3	0	+	0	0	+	+	0	0	+	+
4	+	+	0	0	0	+	0	0	0	+
5	0	0	+	0	0	0	+	+	+	+
6	+	0	+	0	0	0	0	+	0	+
7	0	+	+	0	0	0	0	0	+	+
8	+	+	+	0	0	0	0	0	0	+

$$q_{il}^* = \begin{cases} 1, & \text{if } \exists (a_{i1}, \dots, a_{ik}) : \prod_{k=1}^k a_{ik}^{q_{ik}} = 1 \text{ and } \sum_{k=1}^k 2^{(k-1)} a_{ik} = l. \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Note that if $\mathbf{q}_i = (q_{i1}, \dots, q_{ik}) = (0, \dots, 0)$, then $\eta_{iv} = \prod_{k=1}^k a_{vk}^{q_{ik}} = 1$ holds for all skills vectors in $T(a)$. In that (trivial) case we define $q_i^* = (q_{i1}^*, \dots, q_{iD}^*) = (0, \dots, 0)$.

Table 2 shows an example with DINA skills requirement \mathbf{Q} -matrix rows and the corresponding DINA-equivalent skill requirement vectors.

The definitions of the transformed skill vectors \mathbf{a}^* and the associated q^* are finally used in the linear GDM (von Davier, 2005) model equation. Using $l = \sum_{k=1}^k 2^{(k-1)} a_{vk}$ as introduced above, we obtain

$$P(X_{vi} = 1 | a_v^*, q_i^*) = \frac{\exp(\beta_i + \sum_{n=1}^D \gamma_{in} q_{in}^* a_{vn}^*)}{1 + \exp(\beta_i + \sum_{n=1}^D \gamma_{in} q_{in}^* a_{vn}^*)} = \frac{\exp(\beta_i + \gamma_{il} q_{il}^* a_{vl}^*)}{1 + \exp(\beta_i + \gamma_{il} q_{il}^* a_{vl}^*)}.$$

This holds because by definition there is at most one l with $a_{vl}^* = 1$. If $l = 0$, then all $a_{vn}^* = 0$. In that case, we define $\gamma_{i0} = a_{v0}^* = q_{i0}^* = 0$. If $\mathbf{q} = (0, \dots, 0)$ then also $\mathbf{q}^* = (0, 0, \dots, 0)$. In this case, set $\gamma_{il} = 0$ for all $l = 1, \dots, D$. With these definitions, we have $q_{il}^* a_{vl}^* \in \{0, 1\}$, so we can write

$$P(X_{vi} = 1 | a_v^*, q_i^*) = \left(\frac{\exp(\beta_i)}{1 + \exp(\beta_i)} \right)^{(1 - q_{il}^* a_{vl}^*)} \left(\frac{\exp(\beta_i + \gamma_{il})}{1 + \exp(\beta_i + \gamma_{il})} \right)^{q_{il}^* a_{vl}^*}.$$

By definition of the associated \mathbf{q}^* , \mathbf{a}^* , and l , we have $\eta_{iv} = q_{il}^* a_{vl}^*$ or, if $\mathbf{q} = (0, \dots, 0)$, we have $\eta_{iv} = 1$ for all $a \in T(a)$. Note that we also have $\gamma_{il} = 0$ in that case. Therefore we can write

$$P(X_{vi} = 1 | a_v^*, q_i^*) = \left(\frac{\exp(\beta_i)}{1 + \exp(\beta_i)} \right)^{(1 - \eta_{iv})} \left(\frac{\exp(\beta_i + \gamma_{il})}{1 + \exp(\beta_i + \gamma_{il})} \right)^{\eta_{iv}}.$$

Note that each different \mathbf{a}_v^* may lead to different choices of the DINA-equivalent skill entry index l , that is, $l = f(\mathbf{q}_i^*, \mathbf{a}_v^*)$. Therefore $\gamma_{im} \neq \gamma_{il}$ may hold for different skills vectors

$\mathbf{a}_v^* \neq \mathbf{a}_i^*$. To ensure that the GDM model variant A is equivalent to the DINA model, we introduce the constraint $\gamma_i^* = \gamma_{in}$ for all $n = 1, \dots, D$, and we set $\gamma_i^* = 0$ if the DINA \mathbf{Q} -matrix entry for item i is $\mathbf{q}_i = (0, \dots, 0)$. Then we can write

$$P(X_{vi} = 1 | \mathbf{a}_v^*, \mathbf{q}_i^*) = \left(\frac{\exp(\beta_i)}{1 + \exp(\beta_i)} \right)^{(1-\eta_w)} \left(\frac{\exp(\beta_i + \gamma_i^*)}{1 + \exp(\beta_i + \gamma_i^*)} \right)^{\eta_w}, \quad (4)$$

which enables us to define

$$g_i = \frac{\exp(\beta_i)}{1 + \exp(\beta_i)} \quad (5)$$

and

$$1 - s_i = \frac{\exp(\beta_i + \gamma_i^*)}{1 + \exp(\beta_i + \gamma_i^*)} \quad (6)$$

concluding the proof of the equivalency of GDM variant A and the DINA model.

It is important to reiterate that this equivalency is *based on the linear GDM only, not* on a diagnostic model derived from the general case (1) by involving higher-order skill interactions such as the L-CDM or the logistic G-DINA. This means that the proof presented here is not implied by the definitions of either the L-CDM or the logistic G-DINA.

4. Reparameterization B of the DINA model

The second reparameterization of the DINA model as a linear GDM is based on a slightly different mapping of the DINA skill space and the skill distribution space into a transformed skill space with an associated constrained skill distribution. In order to prepare the mapping into the modified skill space, we introduce a reparameterized vector of required skills based on

$$\mathbf{d} = \sum_{k=1}^k 2^{(k-1)} q_k$$

which establishes a mapping from the set of original skill-requirement vectors $\mathbf{q}_i = (q_{i1}, \dots, q_{ik}) \neq (0, \dots, 0)$ to an integer d . Then, define

$$\mathbf{q}_i^* = (q_{i1}^*, \dots, q_{iD}^*)$$

with $D = 2^k - 1$ and with

Table 3. Skill attribute \mathbf{Q} -matrix for a DINA model with three skills and the equivalent seven-skills \mathbf{Q} -matrix entries for the compensatory GDM

DINA q				Remapped equivalent q^*						
	A	B	C	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0	0	0	0
2	+	0	0	+	0	0	0	0	0	0
3	0	+	0	0	+	0	0	0	0	0
4	+	+	0	0	0	+	0	0	0	0
5	0	0	+	0	0	0	+	0	0	0
6	+	0	+	0	0	0	0	+	0	0
7	0	+	+	0	0	0	0	0	+	0
8	+	+	+	0	0	0	0	0	0	+

$$q_{id}^* = \begin{cases} 1, & \text{for } d = \sum_{k=1}^k 2^{(k-1)} q_k, \\ 0, & j \neq d. \end{cases}$$

If $\mathbf{q}_i = (q_{i1}, \dots, q_{ik}) = (0, \dots, 0)$ then let $q_{ij}^* = 0$ for all $j = 1, \dots, D$.

The reparameterized skills requirement vectors are based on a mapping $g: 2^k \mapsto 2^D$ with $D = 2^k - 1$. This means we have a larger space of potential skill requirements. Note, however, that only 2^k of these actually appear as *existing* skill requirements in the form of a \mathbf{q}^* vector. Table 3 shows an example with three skills; in this case, $D = 2^3 - 1 = 7$.

The skill attribute vectors $\mathbf{a} = (a_1, \dots, a_k)$ are also mapped into reparameterized skill attribute vectors $\mathbf{a}_i^* = (a_{i1}^*, \dots, a_{iD}^*)$. For each possible skill vector $\mathbf{a}_i = (a_{i1}, \dots, a_{ik})$, we define a corresponding transformed skill vector $\mathbf{a}_i^* = (a_{i1}^*, \dots, a_{iD}^*)$ with

$$a_{il}^* = \begin{cases} 1, & \text{if } \exists i : \prod_{k=1}^k a_{ik}^{q_{ik}} = 1 \text{ and } l = \sum_{k=1}^k 2^{(k-1)} q_{ik} \\ 0, & \text{otherwise.} \end{cases}$$

This definition ensures that the transformed skill vector $\mathbf{a}_i^* = (a_{i1}^*, \dots, a_{iD}^*)$ contains skills at all positions that indicate the possession of all skills required (or more skills than required) in the corresponding \mathbf{Q} -matrix vectors $\mathbf{q}_i^* = g(\mathbf{q}_i)$, for $i = 1, \dots, I$. Note that this defines $2^k - 1$ non-zero skill attribute vectors, plus the zero-skills vector. Table 4 shows an example using three skills. We denote the set of these 2^k reparameterized skill vectors by $T(\mathbf{a})$.

It can be seen that the remapped skill attribute patterns do not differ much from DINA skill patterns if there is only one skill involved. Note, however, that as soon as there are two skills present in the DINA skill pattern, the reparameterized skill patterns contain three skills, two for items that require only one of the two (DINA) skills, and one for items that indeed require both (DINA) skills.

The implementation of a model equivalent to the DINA is straightforward given the above definitions. Using the linear GDM (2), we can write

Table 4. Skill attribute vectors for a DINA with three-skills model and the equivalent skill vectors for the compensatory GDM

DINA a				Remapped skill pattern a^*						
	A	B	C	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0	0	0	0
2	+	0	0	+	0	0	0	0	0	0
3	0	+	0	0	+	0	0	0	0	0
4	+	+	0	+	+	+	0	0	0	0
5	0	0	+	0	0	0	+	0	0	0
6	+	0	+	+	0	+	0	+	0	0
7	0	+	+	0	+	+	0	0	+	0
8	+	+	+	+	+	+	+	+	+	+

$$P(X_{vi} = 1 | \mathbf{q}_i^*, \mathbf{a}^*) = \frac{\exp(\beta_i + \sum_k \gamma_{ik} a_k^* q_{ik}^*)}{1 + \exp(\beta_i + \sum_k \gamma_{ik} a_k^* q_{ik}^*)} = \frac{\exp(\beta_i + \gamma_i a_{k(i)}^*)}{1 + \exp(\beta_i + \gamma_i a_{k(i)}^*)},$$

which takes into account that each reparameterized \mathbf{Q} -matrix row q_i^* includes only one non-zero entry. That is, we can write $\gamma_{ik} = \gamma_i$ for all k because there is at most one $q_{ik}^* > 0$. In addition we introduce the constraint

$$\pi_{a'} = 0$$

for all skill vectors \mathbf{a}' not in $T(\mathbf{a})$. That implies there are only $2^k - 1$ parameters required for the estimation of skill vector probabilities, namely only for the elements of $T(\mathbf{a})$. Also, there are only two parameters per item in the reparameterized DINA-equivalent GDM. Each item has a logistic threshold parameter β_i that corresponds to the (not-logistic) DINA guessing parameter via

$$g_i = \frac{\exp(\beta_i)}{1 + \exp(\beta_i)}$$

and the DINA slipping parameter can be expressed as

$$1 - s_i = \frac{\exp(\beta_i + \gamma_i)}{1 + \exp(\beta_i + \gamma_i)}.$$

For a person with skill vector $\mathbf{a}_v = (a_{v1}, \dots, a_{vk})$ and the associated transformed skill vector $\mathbf{a}_v^* = (a_{v1}^*, \dots, a_{vD}^*)$ we have

$$g_i^{1-\eta_{vi}} (1 - s_i)^{\eta_{vi}} = \left(\frac{\exp(\beta_i)}{1 + \exp(\beta_i)} \right)^{1-a_{vk(i)}^*} \left(\frac{\exp(\beta_i + \gamma_i)}{1 + \exp(\beta_i + \gamma_i)} \right)^{a_{vk(i)}^*},$$

which is equivalent to

$$g_i^{1-\eta_{vi}}(1-s_i)^{\eta_{vi}} = \frac{\exp(\beta_i + \gamma_{ik(i)}a_{vk(i)}^*)}{1 + \exp(\beta_i + \gamma_{ik(i)}a_{vk(i)}^*)},$$

and we finally obtain

$$g_i^{1-\eta_{vi}}(1-s_i)^{\eta_{vi}} = \frac{\exp(\beta_i + \sum_k \gamma_{ik(i)}a_{vk(i)}^*q_{ik}^*)}{1 + \exp(\beta_i + \sum_k \gamma_{ik(i)}a_{vk(i)}^*q_{ik}^*)}$$

as the probability of a correct response. Note that the right-hand side is based on the GDM using the transformed skill patterns and \mathbf{Q} -matrix, while the left-hand side is based on the DINA using the original skill and \mathbf{Q} -matrix space. The above result is well defined because there is only one $k \in \{1, \dots, D\}$ with $q_{ik}^* = 1$, so we can write $a_{vk(i)}^* = \sum_{k=1}^D a_{vk}^* q_{ik}^*$. Also $a_{vk(i)}^* = 1 \iff \prod_{k=1}^K a_{vk}^{q_{ik}} = 1$ and $a_{vk(i)}^* = 0$ if $\eta_{vi} = \prod_{k=1}^k a_{vk}^{q_{ik}} = 0$ by definition of the transformed skill vector $\mathbf{a}_v^* = (a_{v1}^*, \dots, a_{vD}^*)$. This proves the equivalency of the DINA model and the second reparameterization as a simple-structure, compensatory GDM.

5. A practical illustration

While Sections 3 and 4 give a mathematical proof of two different model equivalencies showing that there are bijections between the DINA and two variants of a GDM¹ with the \mathbf{Q} -matrices and skill spaces as defined above, this section provides practical evidence for the utility of these results.

For reasons of access and comparability, the data set described by de la Torre (2009) on DINA estimation was used. In addition, in order to eliminate the possibility that one unique implementation of a single program is deemed the cause of agreement between results, the equivalent DINA was estimated with the software mdltm (von Davier, 2005), a program for binary and polytomous GDMs and other latent variable models (von Davier, 2007, von Davier, 2009), while the DINA in its original parameterization was estimated using the Ox (Doornik, 2002) implementation provided by de la Torre (2009).

5.1. Data

The illustration uses a data set with 30 binary items and 1,000 respondents. The \mathbf{Q} -matrix for this data set is described in de la Torre (2009). The data were generated using the DINA model, based on a \mathbf{Q} -matrix containing five skills, which leads to a latent space with $2^5 = 32$ skill profiles. Given that only 30 items are provided, and 10 of these items are pure items (only one skill measured), there are legitimate concerns whether the remaining 20 items that do not cover the remaining $32 - 5$ possible \mathbf{Q} -matrix rows are indeed sufficient to identify all 32 different skill patterns. Equally important, all items that measured more than one skill were unique items in the original \mathbf{Q} -matrix, that is, their required skill pattern was not shared by any other item, thus making a reliable identification of skill patterns less likely.

¹ Obviously, the proven bijections between the DINA and the two equivalent special cases of the GDM also show that these two special cases of the GDM are equivalent models.

Table 5. Q-matrix used in the data example. Instead of showing a matrix with 30 rows, only the unique entries are shown together with the rows in which they appear

	DINA									Equivalent DINA									
In rows	A	B	C	D	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
4,5,9,10,20	0	0	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
3,8	0	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
18,19,30	0	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
2,7	0	1	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
16,17,29	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1
15	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1
27,28	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
1,6	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
13,14,26	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1
12	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1
24,25	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
11	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
22,23	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
21	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

It was therefore decided to work with a reduced skill Q-matrix that contains four skills. These four skills were obtained by collapsing the fourth and fifth skills of the original Q-matrix (de la Torre, 2009) into one. Apart from a more realistic number of latent classes to identify (Haertel, 1989) relative to the number of items, this approach has the added advantage of introducing a slight amount of misfit compared to the original simulated data and thus to test the agreement between the DINA and the equivalent DINA under more realistic conditions. Table 5 shows the four-skill Q-matrix used in this illustration of the model equivalency and the associated equivalent Q-matrix entries defined according to equation (3). In addition, the constraint $\gamma_i^* = \gamma_{ik}\gamma_{ik}$ for all i and $k = 1, \dots, k$ (as defined in Section 3) was used to ensure that the equivalent DINA estimates the same number of parameters as the DINA (see Section 3).

It is apparent that even after the reduction to four skills, many Q-matrix rows represent unique items that require a skill combination not seen in other items. Clearly a Q-matrix with five skills and 30 items would have had almost no replication of item by skill requirement, because there would be 32 possible skill combinations (including no skill being required and all skills being required). This holds unless the average number of skills required per item is close to 1, in which case there would be no difference between compensatory and conjunctive models. Therefore, we focus our analyses on this four-skill Q-matrix. This provides a sufficient number of items with more than one skill requirement, and at least some amount of replications of these requirements in at least two item positions.

5.2. Identifiability of model parameters

The literature on assessing identifiability of diagnostic models is sparse at best. Apart from important caveats with respect to identifiability of diagnostic models discussed by Maris and Bechger (2009) and some general rules of thumb suggested by, among others, Haertel (1989) for constrained latent class models, there is little to be found. Junker and Sijtsma

(2001) only mention indirect indicators of under-identification based on observing slow convergence of the Markov chain Monte Carlo estimation in their discussion of an example of the DINA model, and most other authors do not address identification at all. Huebner (2010), citing von Davier (2005, 2008), states that identification of parameters is increasingly difficult with increasing numbers of skills in the model. While progress has recently been made on Bayesian estimation of semiparametric IRT models (San Martín *et al.*, 2011), little activity can be found in the case of diagnostic classification models. This lack of work, however, is not necessarily limiting explorations of identifiability of a specific diagnostic model (with a specific **Q**-matrix) applied to simulated or real data due to the availability of general criteria to assess local identifiability.

In order to assess the identifiability of model parameters for the DINA-equivalent GDM presented in this paper, we utilize results presented by Goodman (1974) for latent class models. These criteria agree with Catchpole and Morgan's (1997) conditions for lack of parameter redundancy, which is closely related to local identifiability, for a broad class of models. The main result of practical importance is based on a requirement of non-singularity of the information matrix, because assessment of global identifiability is difficult in general families of models (Catchpole & Morgan, 1997; Maris & Bechger, 2009). Formann (1985) notes that, as an alternative to checking the positive definiteness of the information matrix, local identifiability can be checked by assessing whether all eigenvalues of the information matrix are positive. Therefore, in order to assess identification of model parameters for the data example examined with the equivalent DINA model described in the sections above, this criterion was chosen to evaluate the identifiability.

The information matrix also plays an important role in parameter estimation, and in the evaluation of the uncertainty of estimates through its inverse, the variance–covariance matrix of the parameter estimates. For the analysis at hand, the software *mdltm* (von Davier, 2005, 2008) was used to determine an estimate of these matrices after convergence of model parameters. The covariance matrix for the data example is $M(75, 75)$ and contains variances and covariances for each of the $2^4 - 1 = 15$ non-redundant class sizes as well as for the $30 + 30$ (guessing- and slipping-equivalent) item parameters.

An estimate of the standard error for each of the model parameters is provided below. An evaluation of local identifiability of the DINA model parameters (or in this case the equivalent DINA model specified as a special case of the GDM) is not available in the Ox implementation, and to our knowledge has not been undertaken or reported elsewhere in the literature on the DINA, so we focus here on reporting quantities based on the estimation using the DINA-equivalent model.

Table 6 provides an overview of the standard errors of the parameter estimates. The right-hand side of the table contains the standard errors of the 15 non-redundant class sizes, while the left-hand side contains the standard errors of the item parameters. Note that none of the standard errors of class sizes is larger than 0.013, and none of the standard errors of difficulties and slopes is larger than 0.08 and 0.2, respectively. In other words, all standard errors are smaller by a factor of 6–10 than the absolute values of the parameter estimates. This is a reasonable result given that the (DINA and equivalent DINA) model has 75 free parameters and is estimated based on a sample that contains only 1,000 observations.

All eigenvalues are positive, the smallest being 28.03, and the rank of the information matrix is 75, as expected for an identified model with 75 free parameters (Catchpole & Morgan, 1997). The condition number of the information matrix, calculated as the ratio of

Table 6. Standard errors of class sizes and item parameters of the DINA-equivalent GDM for the example data set.

π	$SE \pi$		β	$SE \beta$	γ	$SE \gamma$
0.01998	0.00444	1	-2.43343	0.05506	1.24354	0.05624
0.03444	0.00657	2	-2.39444	0.07018	1.28918	0.05659
0.03486	0.00585	3	-2.37089	0.06496	1.29249	0.05517
0.03613	0.00608	4	-2.33572	0.05816	1.45420	0.06114
0.03841	0.00568	5	-2.26098	0.06453	1.59064	0.07442
0.04024	0.00603	6	-2.22904	0.05904	1.67059	0.08384
0.04337	0.00633	7	-2.20992	0.05760	1.74519	0.10046
0.04569	0.00666	8	-2.18783	0.06356	1.77563	0.08926
0.07682	0.00849	9	-2.13782	0.06481	1.80807	0.08758
0.07877	0.01154	10	-2.12194	0.06290	1.81941	0.09385
0.07998	0.01055	11	-2.08750	0.05649	1.83853	0.13079
0.08451	0.00918	12	-2.07756	0.05427	1.91109	0.06579
0.09011	0.00968	13	-2.07061	0.05545	1.95482	0.12393
0.09270	0.01101	14	-1.97765	0.05593	2.02207	0.06909
0.10083	0.00883	15	-1.83168	0.05472	2.02751	0.04945
		16	-1.78200	0.05365	2.03810	0.06280
		17	-1.73824	0.05339	1.99350	0.13693
		18	-1.71351	0.05425	2.07560	0.03812
		19	-1.68780	0.05449	2.10708	0.14971
		20	-1.65788	0.05084	2.12040	0.09576
		21	-1.42190	0.04567	2.13332	0.03760
		22	-1.33263	0.04407	2.15669	0.12656
		23	-1.31455	0.05065	2.17669	0.03993
		24	-1.29978	0.04225	2.24797	0.04218
		25	-1.30267	0.05103	2.25514	0.06696
		26	-1.24086	0.04941	2.26569	0.03669
		27	-0.58433	0.03796	2.28446	0.03878
		28	-0.54309	0.03622	2.31746	0.19531
		29	-0.52866	0.03892	2.33205	0.06862
		30	-0.37707	0.03944	2.44971	0.09917

the smallest to largest eigenvalue, is $\kappa^{-1} = 5.72 \times 10^{-5}$.² Given the full rank of the matrix and the positivity of all eigenvalues, it can be concluded that the (equivalent) DINA model with the given **Q**-matrix is identifiable for the data set presented here, and that the order of magnitude of standard errors for class size and item parameter estimates is satisfactory given the moderate sample size.

5.3. Model data fit

This section presents a comparison of the log-likelihoods and information criteria obtained under the DINA and equivalent DINA models. Note that the two models were estimated with two different algorithms. As noted above, the DINA model was estimated

² This value appears to be larger than the threshold for concern about identifiability according to, for example, the Mplus user's guide (Muthén and Muthén, 1998–2010, p. 421), which gives 1×10^{-6} as a boundary when evaluating latent class models. Therefore, we will rely mainly on the assessment of the rank and the positivity of all eigenvalues.

Table 7. Model equivalency in terms of log-likelihood and information criteria

	DINA Ox	Equivalent DINA mdltn	Linear GDM
Log-likelihood	-14,632.33	-14,632.33	-14,750.92
Deviance	29,264.66	29,264.66	29,501.85
AIC	29,414.66	29,414.66	29,703.85
BIC	29,782.74	29,782.74	30,199.53
Npar	75	75	101

with the Ox implementation described by de la Torre (2009), while the equivalent DINA model was estimated using the GDM software mdltn (von Davier, 2005). Table 7 shows the results for the DINA, the equivalent DINA specification estimated using the GDM with the **Q**-matrix given on the right-hand side of Table 5, and the linear (compensatory) GDM model using the DINA **Q**-matrix on the right-hand side. The latter was included to provide some level of comparison and to show that for this particular simulated data set, the DINA and equivalent DINA fit somewhat better than alternative models.

Note that the DINA and equivalent DINA produce virtually identical results. Given the mathematical proof in Sections 3 and 4, this should be self-evident, but the empirical illustration provides exemplary numbers that show that the DINA and equivalent DINA produce (modulo potential differences in the third decimal that could not be evaluated because the Ox DINA implementation produces only two decimals in the output) virtually identical results. The comparison with the linear GDM shows that a data set that was simulated using the DINA model can be fitted somewhat better using either of the two DINA variants compared to a compensatory model (here the linear GDM with the four-skill DINA **Q**-matrix). It is important to understand that, while the DINA and equivalent DINA will fit the same data in the same way because the models are mathematically equivalent, they may not (and indeed often in practice will not) fit real data better than a compensatory approach. This particular order was obtained likely because a data set generated with the DINA was used for illustration.

5.4. Equivalency of recovered skill distribution

Table 8 shows the agreement of skill distribution estimates between the DINA and the equivalent DINA model implemented as a constrained GDM. Recall that these estimates were obtained from two separately developed estimation methods, using two different but equivalent models, with completely different interpretations. One is based on a

Table 8. Agreement of skill distribution estimates between DINA and equivalent DINA model implemented as a constrained GDM

Skills	Equivalent DINA	DINA	Difference	Skills	Equivalent DINA	DINA	Difference
0000	0.09271	0.0926	-0.00011	1000	0.09011	0.0901	-0.00001
0001	0.03444	0.0345	0.00006	1001	0.01998	0.0200	0.00002
0010	0.10318	0.1031	-0.00008	1010	0.07682	0.0768	-0.00002
0011	0.04569	0.0457	0.00001	1011	0.04024	0.0402	-0.00004
0100	0.07876	0.0788	0.00004	1100	0.08451	0.0845	-0.00001
0101	0.03612	0.0362	0.00008	1101	0.04337	0.0434	0.00003
0110	0.10083	0.1008	-0.00003	1110	0.07997	0.0008	0.00003
0111	0.03486	0.0349	0.00004	1111	0.03841	0.0384	-0.00001

constrained compensatory diagnostic model with a higher-dimensional – but constrained – skill space, the linear GDM (von Davier & Yamamoto, 2004); the other model is the DINA (Junker & Sijtsma, 2001; Haertel, 1989; Macready & Dayton, 1977), which assumes conjunctive skills.

The level of agreement is extremely high for a sample size of 1,000 observations. The differences between the estimated skill distributions are of the order 10^{-4} or smaller. Note that skill distributions were estimated separately (i.e., there was no borrowing of starting values between implementations; only the uniform distribution of skills was provided as a starting point).

While the mathematical proof of model equivalency should be sufficient in terms of required scientific evidence, Table 8 illustrates using data that there is a very high level of agreement between the numerical values obtained from the DINA and the equivalent DINA models. Note that the Ox DINA output only provides four decimal places here, so the differences in the fifth decimal place could be caused by this truncation. This means that distributions of skills obtained from the two different models are, for all practical purposes, identical.

5.5. Equivalency of recovered slipping and guessing parameters

Table 9 shows the recovered slipping and guessing parameters using equations (5) and (6) together with the corresponding logistic parameters that form the basis of the equivalent DINA model formulated as a constrained GDM. Just as found for the skill distribution, the slipping and guessing parameters obtained from the DINA and the equivalent DINA estimations are virtually identical, as Table 8 shows. As indicated above, this is expected given the mathematical proof of equivalency given in Section 3, but it is reassuring to know that two different numerical algorithms agree on the estimated quantities up to the fourth or fifth decimal place for the DINA and equivalent DINA.

5.6. Agreement of skill classifications

In addition to the three areas of agreement reported above, the agreement of classifications was also examined. For marginal agreement (i.e., component-wise agreement for each of the skills), values of .993, .995, .996, and 1.0 are obtained. It is interesting to note that perfect agreement was obtained for the skill that was a result of collapsing skills 4 and 5 from the original **Q**-matrix used by de la Torre (2009). For complete agreement (i.e. agreement with respect to all four skills simultaneously), a value of .984 is obtained. For agreement on three or more skills, a value of 1.0 is obtained. This means that the skill classifications disagree in only 16 cases on exactly one out of four skills.

Note that these small differences in classifications could be due to the small differences in item and skill distribution parameters, or other numerical differences of implementation. The mdltm software uses double-precision (long) variables, which may be one of the reasons for the small differences in classifications. In any event, the level of agreement between the two implementations is excellent, and in conclusion, there is no empirical evidence of any systematic differences between the estimates obtained from the DINA model and the equivalent DINA model. Table 10 shows these results and also provides agreement measures for the linear (non-equivalent) GDM for comparison.

Note that even the linear (non-equivalent) GDM shows excellent componentwise and very good four-skill agreement. The linear GDM also has an excellent three- or four-skill agreement of .995. This very high agreement between the four-skill linear GDM and the

Table 9. Agreement of DINA and constrained GDM equivalent DINA estimates of slipping and guessing parameters

Item	Slope	Equivalent	Location	Equivalent	DINA g	DINA s	Diff g	Diff s
		DINA g		DINA s				
1	4.5687	0.0882	-2.3356	0.0968	0.0881	0.0968	0.0001	0.0000
2	4.5305	0.0807	-2.4328	0.1093	0.0799	0.1092	0.0008	0.0001
3	4.4958	0.1103	-2.0874	0.0825	0.1102	0.0825	0.0001	0.0000
4	2.9083	0.3579	-0.5843	0.0892	0.3577	0.0894	0.0002	-0.0002
5	2.4870	0.4068	-0.3771	0.1081	0.4067	0.1085	0.0001	-0.0004
6	4.3531	0.0972	-2.2289	0.1068	0.0970	0.1067	0.0002	0.0001
7	4.1509	0.1120	-2.0704	0.1110	0.1117	0.1111	0.0003	-0.0001
8	4.2667	0.1113	-2.0776	0.1007	0.1112	0.1007	0.0001	0.0000
9	2.5783	0.3675	-0.5431	0.1156	0.3673	0.1158	0.0002	-0.0002
10	2.5849	0.3708	-0.5287	0.1134	0.3707	0.1137	0.0001	-0.0003
11	4.6642	0.0944	-2.2610	0.0829	0.0944	0.0832	0.0000	-0.0003
12	4.5102	0.1055	-2.1378	0.0853	0.1055	0.0853	0.0000	0.0000
13	3.3411	0.2137	-1.3027	0.1152	0.2136	0.1153	0.0001	-0.0001
14	3.6388	0.2117	-1.3146	0.0891	0.2117	0.0892	0.0000	-0.0001
15	4.0762	0.1216	-1.9777	0.1092	0.1216	0.1095	0.0000	-0.0003
16	3.5512	0.2243	-1.2408	0.0903	0.2242	0.0907	0.0001	-0.0004
17	3.1811	0.1944	-1.4219	0.1469	0.1943	0.1473	0.0001	-0.0004
18	4.3131	0.2087	-1.3326	0.0483	0.2087	0.0487	0.0000	-0.0004
19	3.6161	0.2142	-1.2998	0.0898	0.2141	0.0901	0.0001	-0.0003
20	4.0550	0.0989	-2.2099	0.1365	0.0988	0.1371	0.0001	-0.0006
21	4.2407	0.1070	-2.1219	0.1073	0.1070	0.1073	0.0000	0.0000
22	3.4903	0.1495	-1.7383	0.1478	0.1495	0.1480	0.0000	-0.0002
23	4.6344	0.1561	-1.6878	0.0499	0.1560	0.0499	0.0001	0.0000
24	3.9095	0.1441	-1.7820	0.1065	0.1440	0.1067	0.0001	-0.0002
25	3.9867	0.1527	-1.7135	0.0934	0.1527	0.0935	0.0000	-0.0001
26	4.8993	0.0836	-2.3944	0.0755	0.0836	0.0760	0.0000	-0.0005
27	4.2139	0.1380	-1.8317	0.0845	0.1380	0.0851	0.0000	-0.0006
28	3.6769	0.1600	-1.6579	0.1172	0.1600	0.1176	0.0000	-0.0004
29	4.0441	0.0854	-2.3709	0.1580	0.0854	0.1589	0.0000	-0.0009
30	3.8221	0.1008	-2.1878	0.1632	0.1007	0.1634	0.0001	-0.0002

Table 10. Agreement between DINA, equivalent DINA, and linear GDM

		Skill 1	Skill 2	Skill 3	Skill 4	Set of 4
Equivalent DINA	Agreement	.993	.996	.995	1.000	.984
	Kappa	.986	.992	.990	1.000	.983
	Cramér's <i>V</i>	.986	.992	.990	1.000	.988
Linear GDM	Agreement	.964	.963	.955	.981	.868

four-skill DINA is another indication of why it may often not be obvious that the DINA and a conjunctive assumption should be adopted. Apart from the very high agreement, the linear GDM may not fit quite as well as the DINA or the equivalent DINA for this data example, but note that the data were simulated using the DINA as the generating model. In any case, the classification agreement and the log-likelihood of $-14,750.92$ still being close to the value for DINA and equivalent DINA indicate that all models examined here fit this DINA generated data set in quite similar ways.

It is important to note here that classifications based on model estimates were compared. The levels of agreement reported here were not obtained based on a comparison to the true (generating) skill profiles, which were not available for analysis, just as is the case in real data analysis. Therefore, the reported values cannot be viewed as indicators of ‘how close to the truth’ the different models are, but rather how well estimates from two different implementations of the DINA and the equivalent DINA agree. Because parameter estimates between DINA and equivalent DINA are virtually identical, the 16 cases in which the two implementations differ in classifications on a single skill can most likely be attributed to numerical differences in the algorithmic implementation of the classifiers.

5.7. Summary and advantages of the proposed approach

The DINA model and the equivalent DINA model, specified as a constrained GDM, yield, for all practical purposes, identical results. While this was to be expected based on the mathematical proof given in Sections 3 and 4, it is useful to know that there are two completely separate estimation methods that show extremely high agreement between estimates. In addition, no parameters were shared between estimations, that is, no starting values other than those used in the standard implementations were used. In addition, the parameters obtained were not constrained or linked in any way before estimation, yielding evidence that each of the estimated parameter vectors is very close to the identified maximum likelihood solution for the data example.

While it may seem to readers who are not fond of mathematical derivations that the results presented are a purely academic exercise, it needs to be pointed out that the results presented here have far-reaching consequences for the interpretation of skills derived from models for cognitive diagnosis. The equivalency result developed here implies that we should be much more reluctant to call a set of skills conjunctive or non-compensatory because models may exist that lead to identical fit of the data based on compensatory skills definitions. This means that another researcher who is less inclined to believe that skills are functioning in a restrictive way assumed in conjunctive models may come up with an alternative model that explains the data equally well, or even better if a less constrained model is chosen.

Model equivalency results should lead researchers to be a little more careful when talking about assumptions and model interpretations such as the conjunctive nature of skills, or, for that matter, the absence or presence of skill hierarchies, or the higher- or lower-level ordering of factors. All these are, in light of the results of research by Bechger *et al.* (2002) – as well as by Maris and Bechger (2004, 2009), Rijmen (2010), Yung *et al.* (1999), and the results presented here – not much more than interpretive categories rather than a reality that provides a unique explanation of the data at hand. As the authors just cited have shown, there are many cases in which there coexist equivalent linear logistic test models, or there may be several equivalent \mathbf{Q} -matrices yielding the same response probabilities for a given diagnostic model, or the testlet IRT model and the higher-order IRT model are nothing but constrained versions of the more general bifactor IRT model. The research presented here adds a different type of model equivalency and shows how a conjunctive model, the DINA, can be re-expressed as a higher-dimensional but constrained compensatory GDM.

It is understandable that a result that may lead to less distinction between models is not appreciated or even rejected (e.g., Bruner & Postman, 1949; Festinger, 1957) by users or proponents of the approach, which has just been shown to be a special case of something

more general. Instead of further discussing this issue, however, we provide some ideas for the practical use of the result presented here for researchers who aim to use the DINA for more complex assessment designs. This list contains estimation and modelling applications opened up by using the equivalent DINA that may be useful beyond the fact that the DINA can be compared to more diagnostic models such as the GDM. More specifically, by using the equivalent DINA in the GDM framework, researchers have, among other things, access to the following:

- (1) Extensions of the (equivalent) DINA to mixture and multilevel general diagnostic models (von Davier, 2010; von Davier, 2007) by using the GDM framework that includes developments of this type. In addition, the software implementation of the GDM also provides item fit (Kunina-Habenicht *et al.*, 2009) and person fit measures for the family of GDMs, now including the equivalent DINA model.
- (2) Multiple-group versions of the (equivalent) DINA model allowing completely separate estimation in one model, or estimation with constraints to link/equate equivalent DINA models across observed populations differing in skill distributions but not in slipping/guessing parameters (von Davier & von Davier, 2007; Xu & von Davier, 2008a).
- (3) Estimation of less constrained versions of the DINA as indicated in Sections 3 and 4 of this paper. As an example, the equivalent DINA definitions include parameter constraints across skills in the **Q**-matrix rows for each item. This ensures that the same number of parameters is estimated in the DINA and the equivalent DINA. One immediate generalization would be to allow two or more different parameters distributed across the equivalent DINA **Q**-matrix. Another generalization would release constraints in the skill distribution in order to relax some of the conjunctive skills assumptions for some items.
- (4) Appropriate use of sampling weights (Xu & von Davier, 2008b). In large-scale applications of psychometric models to population surveys, the sampling design has to be taken into account. One way of doing this is by using sampling weights for parameter estimation and replicate weights for variance estimation (Rutkowski *et al.*, 2010). Resampling-based estimates of standard errors of parameters and skill distributions are provided by the software implementation of the GDM, mdlm (von Davier, 2005; see also Hsieh *et al.*, 2009).
- (5) Assessment of local identifiability of the (equivalent) DINA model and extensions using the methods presented in this section above. This allows an evaluation of whether the model as specified is suitable for the intended purpose of estimating the skill distribution and the slipping and guessing (equivalent) parameters, or whether there is a lack of local identifiability or parameter redundancy is present (Catchpole & Morgan, 1997; Goodman, 1974), so that model-based estimates cannot be interpreted appropriately.

6. Discussion

This paper presents two different reparameterizations of the DINA model as an equivalent, compensatory GDM and an empirical illustration of the mathematical proof using a data set that was simulated using the DINA model (de la Torre, 2009). Both the DINA and the equivalent DINA estimated in the GDM framework contain the same number of parameters but use different skill definitions. What the results presented in this paper mean in terms of interpretability of the DINA model is that there is no clear evidence whether the conjectured conjunctive skill structure is really what produced the data –

even if the DINA seems to fit the data. That is, the example data were fitted in identical ways by the DINA and equivalent DINA, so there is no way to decide which model generated the data. Given the existence of at least two linear (compensatory) DINA-equivalent GDMs, do we really have evidence of the skills needed to solve the items that are conjunctive?

There are two important issues that need to be emphasized at this point. First, the results presented here hold for the DINA model in general; that is, for any \mathbf{Q} -matrix, it was shown how to re-express the skills in the equivalent DINA form. This assertion holds because the general mathematical proof set out in this paper did not make any assumptions about the nature of the \mathbf{Q} -matrix. More specifically, the result is not only valid for simple-structure DINA models, as evident from the proof as well as from the empirical illustration. Second, the results presented here are not based on or implied by models that were recently derived by extending the compensatory GDM. Neither the L-CDM nor the G-DINA served a basis for the equivalencies presented in this paper. Note that extensions of the DINA model to the G-DINA tackle a different issue than what was presented in this paper. In the same way, the L-CDM – derived as an extension of the GDM – does involve skill interactions, and therefore trivially includes the DINA as a special case. The G-DINA (de la Torre, 2011) and other extensions or modifications extend the model space, so that the DINA is a sub-model of a larger framework, but when specified as a DINA, the skills remain conjunctive.

The result presented here is something entirely different. Without extending either the DINA or the compensatory GDM, the result presented here shows the mathematical and empirical equivalence of the DINA and variants A and B of the compensatory GDM by means of alternative skills definitions and constraints. The resulting equivalent DINA has the same parameter count, and identical guessing and slipping parameters can be obtained by back-transformation from the GDM-based equivalent DINA parameters to the DINA model. That means that this paper contains proof that the same data can be explained (fitted) in identical ways using two very different sets of skills and different – conjunctive versus compensatory – modelling approaches.

In consequence, the conjunctive feature of the DINA appears to be only in the eye of the beholder, because at least two equivalent models exist that are based on an alternative (higher-dimensional but constrained) skill space and a simple-structure \mathbf{Q} -matrix for variant B of the equivalencies presented here. Variant A presents a GDM in which each skill vector has only one non-zero entry, while the \mathbf{Q} -matrix is not simple-structure and entries for the items may contain multiple skills requirements. Without constraints on the γ_{ik} (i.e., the GDM's slope parameters), we obtain a more general model than the DINA model. Once we constrain the GDM and assume, for each item, the same γ_i^* for all attribute dimensions $n = 1, \dots, D$ model variant B becomes equivalent to the usual DINA model.

Both equivalency results presented here open avenues for testing the DINA model. In order to ensure equivalence, all but 2^K skill pattern probabilities had to be constrained to 0, and for model variant A, constraints across the GDM slope parameters are required but easily implemented in the GDM software mdltm (von Davier, 2005). If we relax one of these constraints and either allow varying slopes or allow a positive probability for all skill vectors $\mathbf{a}^* \in \{0, 1\}^D$, we obtain a model that is more general than the DINA and can be tested against the DINA-equivalent model in a common modelling framework. This may help evaluate the fit of the DINA model compared to simpler models such as one-dimensional IRT models, which are also special cases of the GDM.

Also, the DINA can be tested by means of relaxing assumptions made in the two variants of the DINA-equivalent GDMs and compared to less restrictive diagnostic models, even though the cautionary notes on the use of diagnostic models in general, voiced for example by Haberman and von Davier (2006) and similarly in Sinharay and Haberman (2008) and von Davier (2009), still apply. In summary, the availability of different DINA-equivalent models can help researchers determine whether the assumption of conjunctive skills in the DINA model is actually appropriate, or whether a much simpler model, such as an IRT model, can fit the data equally well, or whether a diagnostic model with fewer constraints than the DINA can fit the data noticeably better.

References

- Bechger, T. M., Verstralen, H. H. F. M., & Verhelst, N. D. (2002). Equivalent linear logistic test models. *Psychometrika*, 67, 123–136. doi:10.1007/BF02294712
- Box, G. E. P., & Draper, N. R. (1987). *Empirical model building and response surfaces*. New York: Wiley.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153–168. doi:10.1007/BF02294533
- Bruner, J. S., & Postman, L. (1949). On the perception of incongruity: A paradigm. *Journal of Personality*, 18, 206–223. doi:10.1111/j.1467-6494.1949.tb01241.x
- Butter, R., De Boeck, P., & Verhelst, N. (1998). An item response model with internal restrictions on item difficulty. *Psychometrika*, 63, 47–63. doi:10.1007/BF02295436
- Catchpole, E. A., & Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrika*, 84, 187–196. doi:10.1093/biomet/84.1.187
- de la Torre, J. (2009). 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(2), 179–199. doi:10.1007/s11336-011-9207-7
- Doornik, J. A. (2002). *Object-oriented matrix programming using Ox* (Version 3.1) [Computer software]. London, UK: Timberlake Consultants Press.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Formann, A. K. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology*, 38, 87–111. doi:10.1111/j.2044-8317.1985.tb00818.x
- Fu, J., & Li, Y. (2007, April). *An integrative review of cognitively diagnostic psychometric models*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.
- Gibbons, R. D., & Hedeker, D. (1992). Full-information item bifactor analysis. *Psychometrika*, 57, 423–436. doi:10.1007/BF02295430
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215–231. doi:10.1093/biomet/61.2.215
- Haberman, S. J., & von Davier, M. (2006) Some notes on models for cognitively based skills diagnosis. In C. R. Rao & S. Sinharay (Eds.), *Psychometrics, handbook of statistics* Vol. 26. (pp. 1031–1038). Amsterdam, The Netherlands: Elsevier.
- Haertel, E. H. (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, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log linear models with latent variables. *Psychometrika*, 74, 191–210. doi:10.1007/s11336-008-9089-5
- Hsieh, C., Xu, X., & von Davier, M (2009). Variance estimation for NAEP data using a comprehensive resampling-based approach: An application of cognitive diagnostic models. In M. von Davier &

- D. Hastedt (Eds.), *IERI monograph series: Issues and methodologies in large scale assessments*, Vol. 2. (pp. 161–174). Hamburg and Princeton, NJ: IER Institute.
- Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation*, 15(3), 1–7. Retrieved from: <http://pareonline.net/pdf/v15n3.pdf>
- Junker, B. W., & 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, A. A., & 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(2–3), 64–70. doi:10.1016/j.stueduc.2009.10.003
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2, 99–120. doi:10.3102/10769986002002099
- Maris, G., & Bechger, T. M. (2004). Equivalent MIRID models. *Psychometrika*, 69, 627–639. doi:10.1007/BF02289859
- Maris, G., & Bechger, T. (2009). Equivalent diagnostic classification models. *Measurement: Interdisciplinary Research & Perspectives*, 7(1), 41–46. doi:10.1080/15366360802715478
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles: Muthén & Muthén.
- Rijmen, F. (2010). Formal relations and an empirical comparison between the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47, 361–372. doi:10.1111/j.1745-3984.2010.00118.x
- Rijmen, F., & De Boeck, P. (2005). A relation between a between-item multidimensional IRT model and the mixture-Rasch model. *Psychometrika*, 70, 481–496. doi:10.1007/s11336-002-1007-7
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142–151. doi:10.3102/0013189X10363170
- San Martín, E., Jara, A., Rolin, J. M., & Mouchart, M. (2011). On the Bayesian nonparametric generalization of IRT-type models. *Psychometrika*, 76(3), 385–409. doi:10.1007/s11336-011-9213-9
- Sinharay, S., & Haberman, S. J. (2008). How much can we reliably know about what examinees know? *Measurement: Interdisciplinary Research & Perspectives*, 6, 46–49. doi:10.1080/15366360802715486
- Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52, 393–408. doi:10.1007/BF02294363
- von Davier, M. (2005). *A general diagnostic model applied to language testing data*, Research Report RR-05-16. ETS, Princeton, NJ: ETS.
- von Davier, M. (2007). *Mixture general diagnostic models*, Research Report, RR-07-32. Princeton, NJ: ETS.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 287–307. doi:10.1348/000711007X193957
- von Davier, M. (2009). Some notes on the reinvention of latent structure models as diagnostic classification models. *Measurement: Interdisciplinary Research and Perspectives*, 7(1), 67–74. doi:10.1080/15366360902799851
- von Davier, M. (2010). Hierarchical mixtures of diagnostic models. *Psychological Test and Assessment Modeling*, 52(1), 8–28. Retrieved from http://www.psychologie-aktuell.com/fileadmin/download/ptam/1-2010/02_vonDavier.pdf

- von Davier, M., DiBello, L., & Yamamoto, K. (2008) Reporting test outcomes using models for cognitive diagnosis. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 151–176). Toronto, Canada: Hogrefe & Huber.
- von Davier, M., & von Davier, A. (2007). A unified approach to IRT scale linkage and scale transformations. *Methodology*, 3(3), 115–124. doi:10.1027/1614-2241.3.3.115
- von Davier, M., Xu, X., & Carstensen, C. H. (2011). Measuring growth in a longitudinal large scale assessment with a general latent variable model. *Psychometrika*, 76(2), 318–336. doi:10.1007/S11336-011-9202-Z
- von Davier, M., & Yamamoto, K. (2004). *A class of models for cognitive diagnosis*. Paper presented at the 4th Spearman Invitational Conference, ETS, Philadelphia, PA.
- Xu, X., & von Davier, M. (2008a). *Fitting the structured general diagnostic model to NAEP data*. Research Report RR-08-27. ETS, Princeton, NJ: ETS.
- Xu, X., & von Davier, M. (2008b). *Linking with the general diagnostic model*. Research Report, RR-08-08. ETS, Princeton, NJ: ETS.
- Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, 64, 113–128. doi:10.1007/BF02294531

Received 26 October 2012; revised version received 24 October 2012