

ON THE LINK BETWEEN COGNITIVE DIAGNOSTIC MODELS AND KNOWLEDGE SPACE THEORY

JÜRGEN HELLER

UNIVERSITY OF TÜBINGEN

LUCA STEFANUTTI, PASQUALE ANSELMINI AND EGIDIO ROBUSTO

UNIVERSITY OF PADOVA

The present work explores the connections between cognitive diagnostic models (CDM) and knowledge space theory (KST) and shows that these two quite distinct approaches overlap. It is proved that in fact the Multiple Strategy DINA (Deterministic Input Noisy AND-gate) model and the CBLIM, a competence-based extension of the basic local independence model (BLIM), are equivalent. To demonstrate the benefits that arise from integrating the two theoretical perspectives, it is shown that a fairly complete picture on the identifiability of these models emerges by combining results from both camps. The impact of the results is illustrated by an empirical example, and topics for further research are pointed out.

Key words: knowledge structure, latent class model, DINA, DINO, MS-DINA, skill multimap, problem function, skill function, local identifiability.

1. Introduction

The history of science is full of ideas that several people had at the same time, independent of each other. This paper explores the connections between two theories developed in parallel for several decades. The theories at issue are the so-called cognitive diagnostic models (CDM; Bolt, 2007; de la Torre, 2009; DiBello & Stout, 2007; Junker & Sijtsma, 2001; Tatsuoaka, 1990; 2002) and knowledge space theory (KST; Albert, Schrepp, & Held, 1994; Doignon & Falmagne, 1985, 1999; Falmagne, Koppen, Villano, Doignon, & Johanessen, 1990; Falmagne & Doignon, 2011). Although they know of each other, so far there has been surprisingly little communication between the two camps.

Obviously, CDM and KST came from different sides and were developed with slightly different aims. In CDM, the focus was on identifying underlying skills from the very beginning. Referring to the distinction between competence and performance introduced by Chomsky (1965), CDM was interested in the *competence level*. Its ambition was to bring item response theory (IRT) closer to cognitive psychology (Mislevy, 1996) by providing a formal framework for representing cognitive theories. This representation was in terms of skills, which are conceived as discrete cognitive elements that are required to accomplish a task, or to solve a particular problem. On the contrary, KST was first developed as a theory for predicting observable responses to a given collection of problems, thereby operating on the *performance level*. The initial objective was to construct an “efficient machine for the assessment of knowledge” (Doignon & Falmagne, 1999, Preface). Within such a pragmatic perspective, KST was almost exclusively targeting the performance level. Skills came into the picture only later on (Doignon, 1994; Düntsch & Gediga, 1995; Falmagne et al., 1990; Korossy, 1997, 1999).

Correspondence should be made to Jürgen Heller, Department of Psychology, University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany. Email: juergen.heller@uni-tuebingen.de

Due to their distinct starting points, the mathematical approaches taken by the two theories are rather different, too. KST started out as a discrete and deterministic theory for adaptive computerized assessment, exploiting the fact that exact theories need not be numerical. Its fundamental mathematical notions originate from the theories of sets, orders, lattices, and combinatorics. Probabilistic concepts were entered only at a later stage, mainly as a tool to fit KST models to data and to cater for stochastic knowledge assessment. Following and extending the IRT tradition, the CDM approach is probabilistic from the outset, based on latent class models (Lazarsfeld & Henry, 1968). It pays little attention to the discrete mathematics behind the models (like, e.g., the structural relationships among the latent classes, the properties of the relations used to connect subsets of items to subsets of skills, etc.).

Despite all these differences, CDM and KST share important common features: Instead of a dimensional representation based on numerical continua, both approaches implement a discrete, non-numerical representation of individuals. It may thus come as a surprise that these two research strands developed quite independently in parallel. Each of them uses its own notation and defines its own concepts (sometimes attaching different meanings to the same words). The aim of this article is not to provide a complete review but to highlight central notions and methodology developed in both camps, and to point out the benefits of integrating these approaches. Section 2 provides an overview of KST. It introduces its basic concepts and highlights how skills enter the stage and all this can be framed within a probabilistic approach. After introducing important models from CDM in Section 3, the subsequent Section 4 is devoted to a systematic exploration of the links between the two theories. This consideration will reveal substantial overlap by identifying a class of equivalent models. Section 5 demonstrates how theoretical developments in either of the two research strands can benefit from linking them to each other. It characterizes the identifiability of the considered models by drawing upon results on KST models (Heller, 2015; Stefanutti, Heller, Anselmi, & Robusto, 2012) and considerations from CDM (Tatsuoka, 1996, 2002). Finally, the empirical application described in Section 6 points out the practical relevance of the presented results.

2. Knowledge Structures

Doignon and Falmagne (1985) define a *knowledge structure* as a pair (Q, \mathcal{K}) in which Q is a nonempty set (assumed to be finite throughout the paper), and \mathcal{K} is a family of subsets of Q , containing at least Q and the empty set \emptyset . The set Q is called the *domain* of the knowledge structure. Its elements are referred to as *items*¹ and the subsets in the family \mathcal{K} are labeled *knowledge states*. A knowledge state represents the subset of items in the considered domain that an individual masters. More specific structural assumptions on the collection of knowledge states have received particular attention. Among them are the *knowledge spaces* (\mathcal{K} closed under union), the *quasi ordinal knowledge spaces* (\mathcal{K} closed under union and intersection), as well as the *learning spaces*, which are *well-graded* knowledge spaces (i.e., for all knowledge states K there is an item $q \in Q$ such that $K \cup \{q\} \in \mathcal{K}$). In order to describe the link to CDMs, it is necessary to refer to the most general notion of an arbitrary knowledge structure.

The definitions make clear that the theory of knowledge structures was developed with a focus on the solution behavior exhibited on a set of items constituting a knowledge domain. This kind of behavioristic consideration leads to very successful applications, for instance, in educational contexts where there is a curriculum prescribing the content to be covered, allowing for a clear

¹ In KST an item is understood as a type of problem, such as “Draw the line represented by the equation $y = ax + b$, with the coefficients a and b being small integers.” A particular instance of an item would then be “Draw the line represented by the equation $y = 3x - 6$.” The distinction between problem type and instance does not affect the equivalence of CDM and KST models that is established below.

definition of the relevant knowledge domain. There are, however, good reasons not to limit the theory to the kind of operationalism that identifies the state of knowledge with the subset of items an individual is capable of solving. In the sequel, it will be shown that the framework allows for integrating psychological theory by bringing into the picture the underlying cognitive abilities (skills, competencies, ...) responsible for the observable behavior.

2.1. Skills and Knowledge Structures

Falmagne et al. (1990) were the first who sketched how the observed solution behavior may be linked to some underlying cognitive constructs by assigning to each item one or more subsets of skills that are relevant for mastering it. There are several largely independent contributions to this extended framework (Doignon, 1994; Düntsch & Gediga, 1995; Gediga and Düntsch, 2002; Korossy, 1997, 1999). The following will characterize the basics of these skill-based extensions of the theory of knowledge structures.

Let the nonempty finite set Q be a knowledge domain and let S be a nonempty finite set of abstract skills that are assumed to be relevant for solving the items in Q . The main idea formalized below is that identifying the skills that are sufficient for solving each of the considered items provides a complete characterization of the solution behavior. For each subset of skills an individual may be equipped with, the set of items which can be solved within it is uniquely defined. This subset of items constitutes a possible knowledge state, and the collection of all the knowledge states forms a knowledge structure. Details are provided in the sequel.

A *skill multimap* is a triple (Q, S, μ) , where μ is a mapping from Q to 2^{2^S} such that each $\mu(q)$, $q \in Q$, is a nonempty collection of nonempty subsets of S . The elements $C \in \mu(q)$ are called *competencies*. If, additionally, the competencies in each $\mu(q)$ are pairwise incomparable (with respect to set inclusion) then (Q, S, μ) is called a *skill function*. Whenever the basic sets Q and S are clear from the context, we simply use μ to refer to the respective skill multimap, or skill function.

A skill multimap or a skill function may assign more than one competency to an item, representing the fact that there may be more than one set of cognitive operations for solving it. These alternatives can, for example, correspond to different solution paths that may be observed. The skills contained in any of the competencies of a skill multimap are assumed to be sufficient for solving the item. Introducing the notion of a skill function is motivated by the idea that the assigned competencies should even be minimally sufficient for solving the item. Minimality in this sense suggests the above considered property that competencies assigned to an item should be pairwise incomparable. Notice that a skill function can be associated to each skill multimap μ whenever the set S is finite by simply discarding the competencies that are not minimal.

Example 1. Let $Q = \{p, q, r, s\}$ and $S = \{a, b, c\}$. Then the skill multimap defined by $\mu(p) = \{\{a\}\}$, $\mu(q) = \{\{a, b\}, \{b, c\}\}$, $\mu(r) = \{\{b\}, \{c\}\}$, $\mu(s) = \{\{a, b\}, \{c\}\}$ is in fact a skill function. It states that item q can be solved in different ways, requiring the two skills a and b , or the two skills b and c , respectively.

Each skill multimap (Q, S, μ) induces a mapping $p: 2^S \rightarrow 2^Q$ defined by

$$p(T) = \{q \in Q \mid \text{there is a } C \in \mu(q) \text{ such that } C \subseteq T\} \tag{1}$$

for all $T \subseteq S$. We will call p the *problem function induced* by the skill multimap μ . Formally, a *problem function* is defined as a triple (Q, S, p) , where p is a mapping from 2^S to 2^Q that is monotonic with respect to set inclusion, and satisfies $p(\emptyset) = \emptyset$ and $p(S) = Q$.

Given a knowledge domain Q and a skill set S assigning the induced problem function to the corresponding skill function defines a mapping from the set of all skill functions $\mu: Q \rightarrow 2^{2^S}$ to

the set of all problem functions $p: 2^S \rightarrow 2^Q$. Düntsch and Gediga (1995, Proposition 2.3) show that for finite Q and S this mapping actually forms a bijection. Within this context the notions of a skill function and a problem function thus are equivalent.

Consider the following results (see, e.g., Heller & Repitsch, 2008) that partially answer the question in how far properties of skill functions are mirrored by the induced problem functions, and vice versa. Let $\mu: Q \rightarrow 2^{2^S}$ be a skill function and p its induced problem function. Then the following two statements are equivalent.

$$\text{For all } q \in Q \text{ there is } \mu(q) = \{C\} \text{ for some } C \subseteq S, \quad (2)$$

$$p(T_1 \cap T_2) = p(T_1) \cap p(T_2) \text{ for all } T_1, T_2 \subseteq S. \quad (3)$$

Moreover, we also have the equivalence of the two statements

$$\text{For all } q \in Q \text{ each of the competencies } C \in \mu(q) \text{ is a singleton set,} \quad (4)$$

$$p(T_1 \cup T_2) = p(T_1) \cup p(T_2) \text{ for all } T_1, T_2 \subseteq S. \quad (5)$$

A skill function μ is said to be a *conjunctive skill function* if it satisfies (2), and it is called a *disjunctive skill function* if (4) holds. In a conjunctive skill function, for every item there is exactly one subset of skills that is required for solving it. In a disjunctive skill function, on the other hand, a single skill suffices to solve any item, but different skills may allow solution of the same item.

For a given skill function (Q, S, μ) and its induced problem function p , the items in $p(T)$ are exactly those that can be solved within the subset T of skills. Thus, the range of the problem function p forms a knowledge structure (Q, \mathcal{K}) consisting of the knowledge states that are possible given the skill function μ . Notice that the properties of the problem function imply that $\emptyset, Q \in \mathcal{K}$. The knowledge structure (Q, \mathcal{K}) is said to be *delineated* by the skill function μ (Doignon & Falmagne, 1999). In general, a knowledge structure delineated by a skill function is not necessarily closed for union or intersection.

2.2. Probabilistic Framework

A knowledge state represents the subset of items from the considered domain that an individual masters. In general, however, it need not be assumed that a problem is solved if and only if it is mastered. In case of a careless error the item is actually mastered but not solved, while in case of a lucky guess an item is solved without being actually mastered. These types of errors are handled within a probabilistic framework, which is based on dissociating the knowledge state K of a person from the actual given response pattern R . Let $\mathcal{R} = 2^Q$ denote the set of all possible response patterns on the domain Q . Falmagne and Doignon (1988a,b) and Doignon and Falmagne (1999) define a *probabilistic knowledge structure* (Q, \mathcal{K}, P) by specifying a (marginal) distribution $P_{\mathcal{K}}$ on the states of \mathcal{K} , and the conditional probabilities $P(R | K)$ for all $R \in \mathcal{R}$ and $K \in \mathcal{K}$. The marginal distribution $P_{\mathcal{R}}$ on \mathcal{R} then is predicted by

$$P_{\mathcal{R}}(R) = \sum_{K \in \mathcal{K}} P(R | K) \cdot P_{\mathcal{K}}(K). \quad (6)$$

The probabilistic knowledge structure that received most attention is the *basic local independence model* (BLIM), which satisfies the following condition: For each $q \in Q$ there are real constants $0 \leq \beta_q < 1$ and $0 \leq \eta_q < 1$ such that for all $R \in \mathcal{R}$ and $K \in \mathcal{K}$

$$P(R | K) = \left(\prod_{q \in K \setminus R} \beta_q \right) \cdot \left(\prod_{q \in K \cap R} (1 - \beta_q) \right) \cdot \left(\prod_{q \in R \setminus K} \eta_q \right) \cdot \left(\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right). \quad (7)$$

The constants β_q and η_q are interpreted as the probabilities of a careless error and a lucky guess, respectively, on item q .

Skills are easily integrated into this probabilistic framework. For a given skill function (Q, S, μ) that delineates the knowledge structure (Q, \mathcal{K}) consider a probability distribution $P_{\mathcal{C}}$ on the powerset $\mathcal{C} = 2^S$. It specifies the probabilities $P_{\mathcal{C}}(T)$ for all subsets of skills $T \in \mathcal{C}$. Let p be the problem function induced by the skill function μ . Then, with $p^{-1}(\{K\})$ denoting the preimage of the subset $\{K\}$ of \mathcal{K} under p ,

$$P_{\mathcal{K}}(K) = \sum_{T \in p^{-1}(\{K\})} P_{\mathcal{C}}(T) \quad (8)$$

defines a probability distribution on the knowledge states of \mathcal{K} , on which a basic local independence model may be built as outlined above. To be more precise, under the given circumstances a *competence-based local independence model* (CBLIM) satisfies the equations $P_{\mathcal{R}}(R) = \sum_{T \in \mathcal{C}} P(R | T) \cdot P_{\mathcal{C}}(T)$ for all $R \in \mathcal{R}$, and $P(R | T) = P(R | p(T))$ for all $T \in \mathcal{C}$, where the right-hand side of the latter equation is given by (7). The BLIM on $\mathcal{K} = p(\mathcal{C})$ defined by (8) will sometimes be called the BLIM *induced by the CBLIM*. Notice that all this may be easily generalized to conceiving \mathcal{C} as an arbitrary subset of 2^S containing \emptyset and S , for which the notion *competence structure* has been coined (Korossy, 1997, 1999). Sometimes, we will use the term *conjunctive (disjunctive) CBLIM*, whenever the underlying skill function is conjunctive (disjunctive).

3. Cognitive Diagnostic Models

The main purpose of the CDM theory is skill diagnosis, and its main methodological approach is based on a probabilistic modeling of data. A whole range of different models is available (e.g., see DiBello & Stout, 2007; Rupp, Templin, & Henson, 2010), most of them based on the latent class approach (Roussos, Templin, & Henson, 2007).

The notion of a skill (sometimes called *attribute*) is at the heart of the theory. Here, a skill is interpreted as a discrete cognitive component required for accomplishing particular tasks, or for solving a certain class of problems. Skills are conceived as properties of both persons and items. They are formally represented as dichotomous latent variables, as they can either be present or absent in a person, or either required or not required for solving an item. An appropriate set of items is presented to assess the available skills. The cognitive theory behind this procedure specifies the relationship between items and skills, often in form of a binary matrix, called the **Q**-matrix (Tatsuoka, 1990), which essentially assigns to each item a specific subset of skills. Different interpretations of this matrix are viable and give rise to different classes of cognitive diagnostic models, like the important classes of noncompensatory and compensatory models (Rupp et al., 2010). In noncompensatory models all skills assigned to a problem are necessary for solving it, while in compensatory models the absence of some skills can be compensated by the presence of other skills.

For an overview of the most important models belonging to the two classes, the reader is referred to Rupp et al. (2010). To mention some of them, well-known noncompensatory models are the Deterministic Inputs Noisy AND-gate model (DINA; Haertel, 1984, 1989, 1990), the

Noisy Input Deterministic AND-gate model (NIDA; Junker & Sijtsma, 2001), and the conjunctive Multiple Classifications Latent Class Model (MCLCM, Maris, 1999). The class of compensatory models comprises, among others, the disjunctive and the compensatory MCLCM (Maris, 1999) and the Deterministic Input Noisy OR-gate model (DINO; Templin & Henson, 2006). More recent developments provide generalizations of these models, among them the MS-DINA model (de la Torre & Douglas, 2008), the log-linear general diagnosis model (LCDM; Henson, Templin, & Willse, 2009), the generalized DINA model (GDINA; de la Torre, 2011), and the general diagnostic model (von Davier, 2005). This article focusses mainly on the DINA, DINO, and MS-DINA models, as they will be shown to be in direct correspondence to the probabilistic KST models exposed above. The concluding section, however, will take some first steps toward generalizing the presented approach in order to capture models like the GDINA or the LCDM.

3.1. The DINA Model

The DINA model is a noncompensatory CDM implementing a conjunctive rule. The relationship between items and skills is many-to-many, in the sense that a single item might be related to more than one skill, and every single skill might be relevant to more than one item. This kind of relationship is formally expressed by a binary matrix, called the \mathbf{Q} -matrix, with rows corresponding to items and as columns to skills. A single entry \mathbf{Q}_{jk} of this matrix equals 1 if skill k is required for solving item j , and is 0 otherwise².

A second key concept is that of a *knowledge state*, a binary vector that represents the set of skills possessed by some individual. Notice that this meaning is different from that of a knowledge state in KST. We will resolve this clash of terminology below (see Section 4). For a single skill k , latent variables $\alpha_{ik} \in \{0, 1\}$ are introduced to indicate that individual i has available skill k ($\alpha_{ik} = 1$), or not ($\alpha_{ik} = 0$). If n denotes the total number of skills then the knowledge state of individual i is the binary vector $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in})$.

A third element in the model is the observed response $x_{ij} \in \{0, 1\}$ of an individual i to an item j , where 0 means that the response is wrong, and 1 that it is correct. The *response pattern* of individual i then is a binary vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$, where m is the total number of items in the test.

Given these three elements, the DINA model assumes the existence of a deterministic latent response ξ_{ij} of individual i to item j . Such a latent response is labeled “correct” ($\xi_{ij} = 1$) if individual i possesses *all* the skills that are required by item j , and “wrong” ($\xi_{ij} = 0$) in all other cases. Given an $m \times n$ \mathbf{Q} -matrix and a knowledge state $\boldsymbol{\alpha}_i$, this is formally captured by the equation

$$\xi_{ij} = \prod_{k=1}^n \alpha_{ik}^{\mathbf{Q}_{jk}}, \quad (9)$$

adopting the convention that $0^0 = 1$. The latent vector $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{im})$ is called the *ideal response pattern* of person i . This means that, in absence of noise, the equality $\mathbf{x}_i = \boldsymbol{\xi}_i$ holds true.

The probabilistic part of the model takes into account that noise is likely to occur with real respondents. This will deteriorate the ideal response pattern to some extent, so that in general $\mathbf{x}_i \neq \boldsymbol{\xi}_i$. For each of the items in the test two sources of noise are considered: *slip* and *guessing*. The slip parameter s_j is regarded as the conditional probability of failing item j given a respondent has available all skills required by it, whereas the guessing parameter g_j is regarded as the conditional probability of solving j given that not all required skills are available.

² Notice that this notation deviates from the standard in CDM (using Q and q_{jk} instead) in order to avoid a notational clash with knowledge domain Q and item q .

The probabilistic link between the unobservable ideal response pattern ξ_i of a person, and the corresponding observed response pattern \mathbf{x}_i is obtained by assuming local stochastic independence among the responses to the items, given the knowledge state of a person. Let \mathbf{X}_i denote a random vector whose realizations are the response patterns \mathbf{x}_i . Under local independence the conditional probability that individual i exhibits response pattern \mathbf{x}_i , given that her knowledge state is α_i , takes on the form (see, e.g., Junker & Sijtsma, 2001; de la Torre, 2009)

$$P(\mathbf{X}_i = \mathbf{x}_i | \alpha_i) = \prod_{j=1}^m [(1 - s_j)^{x_{ij}} s_j^{1-x_{ij}}]^{\xi_{ij}} [g_j^{x_{ij}} (1 - g_j)^{1-x_{ij}}]^{1-\xi_{ij}}. \quad (10)$$

This equation, together with a (to be estimated) probability distribution $P(\alpha_l)$ on the 2^n theoretically possible binary vectors α_l , predicts the marginal probability of response pattern \mathbf{x}_i by

$$P(\mathbf{X}_i = \mathbf{x}_i) = \sum_{l=1}^{2^n} P(\mathbf{X}_i = \mathbf{x}_i | \alpha_l) P(\alpha_l).$$

3.2. The DINO Model

The DINO model is the disjunctive counterpart of the DINA. This compensatory model is based on the assumption that each of the skills assigned to an item by the \mathbf{Q} -matrix is sufficient for solving it. This assumption is formalized by the following deterministic relation between knowledge states α_i and ideal response patterns ξ_i :

$$\xi_{ij} = 1 - \prod_{k=1}^n (1 - \alpha_{ik})^{Q_{jk}}. \quad (11)$$

The probabilistic (noisy) part of the DINO is identical to that of the DINA as specified by (10).

3.3. The Multiple Strategy DINA Model

In concrete applications an item can allow for more than one solution strategy (see, for example, the fraction subtraction problems discussed in Section 6 below). Not so in the DINA model, where for each item there is a unique minimal subset of skills permitting its solution. The multiple strategy DINA model is an attempt to overcome this limitation.

In the MS-DINA model a number M of \mathbf{Q} -matrices \mathbf{Q}_v , $v = 1, 2, \dots, M$, is given, all defined on the same sets Q and S . All matrices have the same number of rows, representing items, but may have different numbers of columns, representing skills. For any given row (item) j , every \mathbf{Q} -matrix \mathbf{Q}_v specifies an alternative subset of skills that are minimally sufficient for solving the same item j . Thus the different \mathbf{Q} -matrices can be viewed as representing different strategies for solving the items represented by their rows.

For every \mathbf{Q} -matrix \mathbf{Q}_v , a (partial) ideal response pattern $\xi_{vi} = \{\xi_{vij}\}$ is defined, whose single elements are obtained by the conjunctive rule

$$\xi_{vij} = \prod_{k=1}^n \alpha_{ik}^{Q_{vjk}},$$

where $\alpha_i = \{\alpha_{ik}\}$ is a knowledge state. This rule is essentially identical to the DINA rule. Then, the overall ideal response pattern ξ_i corresponding to α_i is composed by taking

$$\xi_{ij} = \max\{\xi_{1ij}, \xi_{2ij}, \dots, \xi_{Mij}\}.$$

For convenience, we define the maximum of two vectors of equal size elementwise as the vector $\mathbf{z} = \max\{\mathbf{x}, \mathbf{y}\}$, with each component z_i being the maximum of x_i and y_i . The extension of this operation to an arbitrary number of vectors is straightforward. With this notation, the overall performance state corresponds to $\xi_i = \max\{\xi_{1i}, \xi_{2i}, \dots, \xi_{Mi}\}$. As in the DINA and DINO models, the noisy part in the MS-DINA model is specified by (10).

4. Linking DINA, MS-DINA and DINO to CBLIM

This section shows that there exists a direct correspondence between the MS-DINA model and the competence-based extension of the BLIM, which was called a CBLIM. Although the formulation of the CBLIM relies on set theoretical notions, and the DINA, DINO, and MS-DINA models are based on vectorial representations, the underlying concepts are shown to coincide.

In order to spell out this correspondence, notice that for any finite set $A = \{a_1, a_2, \dots, a_l\}$, $l > 0$, there exists a bijection $\{0, 1\}^l \rightarrow 2^A$ that establishes a one-to-one link between binary vectors $\mathbf{z} \in \{0, 1\}^l$ and subsets $B \subseteq A$ by the equivalence $z_i = 1$ if and only if $a_i \in B$. If this equivalence holds, we will occasionally say that B is the *set representation* for \mathbf{z} . For $A = \{a_1, a_2, a_3, a_4, a_5\}$ the subsets $\{a_1, a_2, a_4\}$, $\{a_2, a_3, a_5\}$ are set representations for the vectors $(1, 1, 0, 1, 0)$, $(0, 1, 1, 0, 1)$, for instance.

This bijection may now be applied to the basic entities of the MS-DINA model, which are the observed response pattern \mathbf{x}_i , the ideal response pattern ξ_i , and the knowledge state α_i of individual i . Let $Q = \{q_1, q_2, \dots, q_m\}$ be a finite set of items and $S = \{s_1, s_2, \dots, s_n\}$ a finite set of abstract skills. Then set representations for \mathbf{x}_i and ξ_i are obtained through functions ρ and κ from $\{0, 1\}^m$ to 2^Q defined by

$$\begin{aligned}\rho(\mathbf{x}_i) &= \{q_j \in Q : x_{ij} = 1\}, \\ \kappa(\xi_i) &= \{q_j \in Q : \xi_{ij} = 1\},\end{aligned}$$

and for α_i via the function $\sigma : \{0, 1\}^n \rightarrow 2^S$ defined by

$$\sigma(\alpha_i) = \{s_k \in S : \alpha_{ik} = 1\}.$$

These definitions formally link the binary vector representation of the MS-DINA model to the set representation used in the CBLIM. There is, however, a clash of terminology that we need to take care of. Both approaches refer to \mathbf{x}_i and $\rho(\mathbf{x}_i)$, respectively, as a response pattern, but they attach different meanings to the term knowledge state. While in CDM this notion is located at the competence level and captures the skills available to an individual, in KST it is located at the performance level and captures the items that the individual is capable of solving. To avoid ambiguities, henceforth we will refer to α_i and $\rho(\alpha_i)$, respectively, as a *competence state*, and to ξ_i and $\kappa(\xi_i)$, respectively as a *performance state*. For ease of reference this terminology is summarized in Table 1.

We may turn now to the fundamental link between the two models, which involves a \mathbf{Q} -matrix collection on the side of the MS-DINA model, and a *skill function* (or a skill multimap) on the

TABLE 1.
Corresponding notions and terminology used.

CDM		KST	Term used
x_i	$\xrightarrow{\rho}$	R	Response pattern
ξ_i	$\xrightarrow{\kappa}$	K	Performance state
α_i	$\xrightarrow{\sigma}$	C	Competence state

side of the CBLIM. A collection $\mathcal{M} = \{\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_M\}$ of $|\mathcal{Q}| \times |S|$ binary matrices, where \mathcal{Q} is a set of items and S is a set of skills is called a **Q-matrix collection** on the sets \mathcal{Q} and S , if every matrix $\mathbf{Q}_m \in \mathcal{M}$ satisfies the following condition: for every row i there is a column j such that $\mathbf{Q}_{mij} = 1$. For each $\mathbf{Q}_v \in \mathcal{M}$ we may then define

$$\mu_v(q_j) = \{\{s_k \in S : \mathbf{Q}_{vjk} = 1\}\} \quad (12)$$

for all $q_j \in \mathcal{Q}$, which provides the *conjunctive skill function* (\mathcal{Q}, S, μ_v) corresponding to \mathbf{Q}_v . Conversely, given a conjunctive skill function μ_v then

$$\mathbf{Q}_{vjk} = 1 \iff s_k \in C \text{ for } C \in \mu_v(q_j) \quad (13)$$

defines a **Q-matrix** \mathbf{Q}_v . Moreover, a skill multimap $\mu : \mathcal{Q} \rightarrow 2^{2^S}$ can be established by requiring

$$\mu(q) = \bigcup_{v=1}^{|\mathcal{M}|} \mu_v(q) \quad (14)$$

for every item $q \in \mathcal{Q}$ which will be called the *skill multimap corresponding to the Q-matrix collection* \mathcal{M} .

By (14) it is always possible to construct a skill multimap, and thus a skill function (see Section 2.1) corresponding to some **Q-matrix collection** \mathcal{M} . The next statement asserts that the converse operation is always possible, too.

Proposition 1. *Every skill function (\mathcal{Q}, S, μ) corresponds to some Q-matrix collection \mathcal{M} on \mathcal{Q} and S .*

Proof. The statement is proven by construction. For every item $q \in \mathcal{Q}$, let the subsets in $\mu(q)$ be arbitrarily ordered, and let C_{ql} denote the l -th subset. For $M = \max\{|\mu(q)| : q \in \mathcal{Q}\}$ and $l = 1, 2, \dots, M$, define the l -th conjunctive skill function μ_l by requiring $\mu_l(q) = \{C_{qr}\}$ for every $q \in \mathcal{Q}$, with $r = \min\{l, |\mu(q)|\}$. That is, for $q \in \mathcal{Q}$ and $\mu(q) = \{C_{q1}, C_{q2}, \dots, C_{qm}\}$, $m \leq M$, we require that for any $l = 1, 2, \dots, M$,

$$\mu_l(q) = \begin{cases} \{C_{ql}\} & \text{if } l < m, \\ \{C_{qm}\} & \text{if } l \geq m. \end{cases}$$

It is clear that $\mu_l(q) = \mu_m(q)$ for all $l \in \{m, m + 1, \dots, M\}$ and hence $\bigcup_{l=1}^M \mu_l(q) = \bigcup_{l=1}^m \{C_{ql}\} = \mu(q)$. Then the result follows because for each of the conjunctive skill functions μ_l there exists a unique **Q-matrix** \mathbf{Q}_l , which is defined via (13). \square

TABLE 2.
Q-matrices for Example 2.

Item	Skills														
	Q ₁			Q ₂			Q ₃			Q ₄			Q ₅		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
p	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
q	0	1	1	1	1	0	1	1	0	0	1	1	1	1	0
r	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0
s	1	1	0	0	0	1	0	0	1	1	1	0	0	0	1

The relationship between Q-matrix collections and skill functions is many-to-one. The following example shows that the skill function corresponding to a given collection of Q-matrices \mathcal{M} is unique, while there may be more than one collection that induces the same skill function.

Example 2. For $Q = \{p, q, r, s\}$ and $S = \{a, b, c\}$, consider the two collections $\mathcal{M}_1 = \{Q_1, Q_2\}$ and $\mathcal{M}_2 = \{Q_3, Q_4, Q_5\}$, where the Q-matrices are defined according to Table 2. With items and skills representing rows and columns as indicated, it is routine to check that both \mathcal{M}_1 and \mathcal{M}_2 induce the same skill function defined by $\mu(p) = \{\{a\}\}$, $\mu(q) = \{\{a, b\}, \{b, c\}\}$, $\mu(r) = \{\{b\}, \{c\}\}$, $\mu(s) = \{\{a, b\}, \{c\}\}$, already considered in Example 1.

Having clarified the relation between Q-matrix collections and skill functions, we can now characterize the problem function corresponding to the latter. The following is immediate from the definitions (see also Heller & Repitsch, 2008, Lemma 4).

Proposition 2. Let $\mathcal{M} = \{Q_1, \dots, Q_M\}$ be a Q-matrix collection on Q and S , and let μ be the skill function induced by \mathcal{M} , and p be the corresponding problem function. Then, for every subset $T \subseteq S$,

$$p(T) = \bigcup_{v=1}^{|\mathcal{M}|} p_v(T),$$

where p_v is the problem function corresponding to the conjunctive skill function μ_v of the matrix $Q_v \in \mathcal{M}$.

Notice that in the following main result of this section, which establishes a link between the MS-DINA and the CBLIM, the delineated performance structure is of the most general form. In particular, it need not be closed under intersection or union, and is not necessarily well-graded.

Proposition 3. Let \mathcal{M} be a Q-matrix collection on Q and S , and μ be the corresponding skill function. Moreover, let α_i be a competence state, and $\xi_i = \max_{v=1}^{|\mathcal{M}|} \{\xi_{vi}\}$, with

$$\xi_{vij} = \prod_{k=1}^n \alpha_{ik}^{Q_{vjk}},$$

be the corresponding performance state. Then the problem function p corresponding to μ is such that

$$p \circ \sigma(\alpha_i) = \kappa(\xi_i).$$

Proof. We first show the equation for the problem function p_v corresponding to the conjunctive skill functions μ_v , which is induced by any of the \mathbf{Q} -matrices \mathbf{Q}_v from \mathcal{M} .

By (9), $\xi_{ij} = 1$ holds true iff the implication $\mathbf{Q}_{vjk} = 1 \implies \alpha_{ik} = 1$ holds for all $k = 1, 2, \dots, n$. Since μ_v is conjunctive, let $\mu_v(q_j) = \{C\}$. Then, by (13), $\mathbf{Q}_{vjk} = 1$ iff $s_k \in C$. Moreover, by the definition of σ , $\alpha_{ik} = 1$ iff $s_k \in \sigma(\alpha_i)$. Furthermore, by the definition of the function κ , $\xi_{ij} = 1$ iff $q_j \in \kappa(\xi_i)$. Thus we have that $q_j \in \kappa(\xi_i)$ iff the implication $s_k \in C \implies s_k \in \sigma(\alpha_i)$ holds for all $s_k \in S$, which in turn holds iff $C \subseteq \sigma(\alpha_i)$ iff $q_j \in p_v \circ \sigma(\alpha_i)$. Hence $\kappa(\xi_i) = p_v \circ \sigma(\alpha_i)$.

Due to this result and Proposition 2 we can write:

$$p \circ \sigma(\alpha_i) = \bigcup_{v=1}^{|\mathcal{M}|} p_v \circ \sigma(\alpha_i) = \bigcup_{v=1}^{|\mathcal{M}|} \kappa(\xi_{vi}).$$

Then, given any two binary vectors \mathbf{x} and \mathbf{y} of the same size, the equation $\kappa(\max\{\mathbf{x}, \mathbf{y}\}) = \kappa(\mathbf{x}) \cup \kappa(\mathbf{y})$ for their elementwise maximum is easily verified. Thus we have the equality

$$\bigcup_{v=1}^{|\mathcal{M}|} \kappa(\xi_{vi}) = \kappa\left(\max_{v=1}^{|\mathcal{M}|} \{\xi_{vi}\}\right) = \kappa(\xi_i)$$

which completes the proof. □

The next step, in examining correspondences between the MS-DINA model and the CBLIM, is to show that the central local independence equation (10) of the DINA, DINO and MS-DINA models is, indeed, equivalent to the local independence equation (7) of the respective BLIM. In the first place, let the equalities $\beta_{q_j} = s_j$ and $\eta_{q_j} = g_j$ hold true for all items $q_j \in \mathcal{Q}$. Given this, we observe that, for any item $q_j \in \mathcal{Q}$, any response pattern $\mathbf{x}_i \in \{0, 1\}^m$ and any performance state $\xi_i \in \{0, 1\}^m$ the following four equivalences hold true

1. $\xi_{ij}x_{ij} = 1 \iff q_j \in \rho(\xi_i) \cap \rho(\mathbf{x}_i)$,
2. $\xi_{ij}(1 - x_{ij}) = 1 \iff q_j \in \rho(\xi_i) \setminus \rho(\mathbf{x}_i)$,
3. $(1 - \xi_{ij})x_{ij} = 1 \iff q_j \in \rho(\mathbf{x}_i) \setminus \rho(\xi_i)$,
4. $(1 - \xi_{ij})(1 - x_{ij}) = 1 \iff q_j \in \mathcal{Q} \setminus [\rho(\mathbf{x}_i) \cup \rho(\xi_i)]$.

So, indeed, one can write

$$\begin{aligned} P(\mathbf{X}_i = \mathbf{x}_i | \alpha_i) &= \prod_{j=1}^m [(1 - s_j)^{x_{ij}} s_j^{1-x_{ij}}]^{\xi_{ij}} [g_j^{x_{ij}} (1 - g_j)^{1-x_{ij}}]^{1-\xi_{ij}} \\ &= \prod_{j=1}^m (1 - s_j)^{x_{ij}\xi_{ij}} \prod_{j=1}^m s_j^{(1-x_{ij})\xi_{ij}} \prod_{j=1}^m g_j^{x_{ij}(1-\xi_{ij})} \prod_{j=1}^m (1 - g_j)^{(1-x_{ij})(1-\xi_{ij})} \\ &= \prod_{q_j \in \rho(\mathbf{x}_i) \cap \rho(\xi_i)} (1 - \beta_{q_j}) \prod_{q_j \in \rho(\xi_i) \setminus \rho(\mathbf{x}_i)} \beta_{q_j} \prod_{q_j \in \rho(\mathbf{x}_i) \setminus \rho(\xi_i)} \eta_{q_j} \prod_{q_j \in \mathcal{Q} \setminus [\rho(\mathbf{x}_i) \cup \rho(\xi_i)]} (1 - \eta_{q_j}). \end{aligned}$$

Substituting R for $\rho(\mathbf{x}_i)$ and K for $\rho(\xi_i)$ according to Table 1 finally provides the local independence equation (7) of the BLIM.

Obviously, the DINA model is a special case of the CBLIM. This is also true for the DINO model, which arises when the skill function μ is disjunctive, rather than conjunctive. In the first

place we note that any conjunctive skill function (Q, S, μ_c) can be turned into a disjunctive skill function (Q, S, μ_d) by defining $\mu_d(q) = \{\{s\} : s \in \mu_c(q)\}$ for all items $q \in Q$. In the second place we observe the following trivial relationship between a \mathbf{Q} -matrix and a disjunctive skill function.

Proposition 4. Consider the mapping $\mu : Q \rightarrow 2^{2^S}$ such that, for all $q_j \in Q$,

$$\mu(q_j) = \{\{s_k\} : s_k \in S, \mathbf{Q}_{jk} = 1\}.$$

Then μ is a disjunctive skill function if and only if, for every item $q_j \in Q$, there is a skill s_k such that $\mathbf{Q}_{jk} = 1$.

When Proposition 4 holds for a skill function μ and a \mathbf{Q} -matrix \mathbf{Q} , we say that μ is the disjunctive skill function corresponding to \mathbf{Q} .

Doignon and Falgagne (1999, Theorem 4.14) establish an intimate relationship between conjunctive and disjunctive skill functions. This result is restated here, and adapted to the case in which a conjunctive and a disjunctive skill function correspond to the same \mathbf{Q} -matrix.

Proposition 5. Given a \mathbf{Q} -matrix \mathbf{Q} , let μ_c and μ_d be, respectively, the conjunctive and the disjunctive skill functions corresponding to \mathbf{Q} . Then the performance structures delineated by μ_c and μ_d are dual one another. In particular, for any $T \subseteq S$,

$$p_c(S \setminus T) = Q \setminus p_d(T),$$

where p_c and p_d are the problem functions corresponding to μ_c and μ_d , respectively.

The following proposition provides the link between the DINO and the CBLIM.

Proposition 6. Let (Q, S, μ_d) be a disjunctive skill map corresponding to a \mathbf{Q} -matrix \mathbf{Q} , α_i be a competence state, and ξ_i be the performance state whose components are given by (11). Then, indicating with p_d the problem function corresponding to μ_d ,

$$p_d \circ \sigma(\alpha_i) = \kappa(\xi_i).$$

Proof. Let μ_c be the conjunctive skill function corresponding to \mathbf{Q} , and p_c be the corresponding problem function. By Proposition 5, and recalling that both κ and σ are bijections, we can write $p_d \circ \sigma(\alpha_i) = Q \setminus p_c(S \setminus \sigma(\alpha_i)) = \kappa \circ \kappa^{-1}(Q \setminus p_c \circ \sigma(1 - \alpha_i)) = \kappa(1 - \kappa^{-1} \circ p_c \circ \sigma(1 - \alpha_i))$. Then we observe that α_i and ξ_i satisfy (11) iff $\tilde{\alpha}_i = 1 - \alpha_i$ and $\tilde{\xi}_i = 1 - \xi_i$ satisfy (9). Therefore the equality $p_c \circ \sigma(\tilde{\alpha}_i) = \kappa(\tilde{\xi}_i)$ holds true. Thus $p_d \circ \sigma(\alpha_i) = \kappa(1 - \kappa^{-1} \circ p_c \circ \sigma(\tilde{\alpha}_i)) = \kappa(1 - \kappa^{-1} \circ \kappa(\tilde{\xi}_i)) = \kappa(\xi_i)$. \square

5. Identifiability

Identifiability of a parametric model guarantees that no two different sets of parameter values lead to exactly the same prediction. Only models satisfying this property allow for uniquely determining the parameters in an application. This section will not provide results on how to determine the identifiability of the BLIM (for this see Heller, 2015; Stefanutti et al., 2012). It rather intends to explore the implications of the intimate relationship between the CBLIM and its induced BLIM for the identifiability of the former. For doing this, it refers to a general framework

to treat (local) identifiability of parametric models introduced by [Bamber and van Santen \(1985, 2000\)](#). Within this framework, a model is regarded as a triple (Θ, f, Φ) , where $\Theta \subseteq \mathbb{R}^m$ is called the model's *parameter space*, $\Phi \subseteq \mathbb{R}^n$ is the model's *outcome space*, and $f: \Theta \rightarrow \Phi$ is the so-called *prediction function* of the model. The model's prediction $f(\theta)$ for a given parameter vector $\theta \in \Theta$ provides an outcome in Φ . Then a model (Θ, f, Φ) is *identifiable* if its prediction function f is one-to-one, and it is *locally identifiable* at a given point $\theta_0 \in \Theta$ if f is one-to-one when restricted to points within some distance $\epsilon > 0$ from θ_0 .

5.1. Theoretical Results

For a given BLIM define the parameter space $\Theta_{\mathcal{K}}$ as follows. Let $\beta = (\beta_q)_{q \in Q}$ and $\eta = (\eta_q)_{q \in Q}$ denote the parameter vectors of the item-specific careless error and guessing probabilities, respectively, and let $\pi = (\pi_K)_{K \in \mathcal{K}^*}$ with $\mathcal{K}^* = \mathcal{K} \setminus \{Q\}$ denote the parameter vector of independent state probabilities $\pi_K = P_{\mathcal{K}}(K)$, $K \in \mathcal{K}^*$. Then the parameter vectors $\theta_{\mathcal{K}} = (\beta, \eta, \pi)$ consist of $m_{\mathcal{K}} = 2 \cdot |Q| + |\mathcal{K}| - 1$ components, all of which are assumed to be nonzero. Moreover, the following constraints apply:

$$\sum_{L \in \mathcal{K}^*} \pi_L < 1, \tag{C1}$$

$$\beta_q + \eta_q < 1 \text{ for all } q \in Q. \tag{C2}$$

Restriction (C2) means nothing else but that a correct response is more likely if the item is mastered, than if it is not mastered. Equivalently, an incorrect response is more likely if the item is not mastered, than if it is mastered. These relations are at the core of the notion of a performance state (knowledge state in KST), and form the essence of any stochastic procedure for knowledge assessment that intends to uncover the performance state of an individual given the observed responses ([Heller and Repitsch, 2012](#)). The parameter space $\Theta_{\mathcal{K}}$ then is defined by

$$\Theta_{\mathcal{K}} = \{\theta_{\mathcal{K}} \in (0, 1)^{m_{\mathcal{K}}} \mid \theta_{\mathcal{K}} \text{ satisfies (C1) and (C2)}\}.$$

Let f denote the prediction function of the BLIM defined on $\Theta_{\mathcal{K}}$.

Considering a CBLIM, we are given an abstract set S of skills via a skill function (Q, S, μ) and a probability distribution $P_{\mathcal{C}}$ on a competence structure \mathcal{C} (e.g., on the powerset 2^S), which is captured by parameters $\pi_{\mathcal{C}} = P_{\mathcal{C}}(C)$, with $C \in \mathcal{C}^* = 2^S \setminus \{S\}$ denoting the competence states. Again, it is assumed that the state probabilities are nonzero, and that they satisfy

$$\sum_{T \in \mathcal{C}^*} \pi_T < 1. \tag{C3}$$

Then the vectors $\theta_{\mathcal{C}} = (\beta, \eta, \pi = (\pi_C)_{C \in \mathcal{C}^*})$ have $m_{\mathcal{C}} = 2 \cdot |Q| + |\mathcal{C}| - 1$ components and the parameter space $\Theta_{\mathcal{C}}$ is defined by

$$\Theta_{\mathcal{C}} = \{\theta_{\mathcal{C}} \in (0, 1)^{m_{\mathcal{C}}} \mid \theta_{\mathcal{C}} \text{ satisfies (C2) and (C3)}\}.$$

Let g denote the function mapping $\Theta_{\mathcal{C}}$ onto $\Theta_{\mathcal{K}}$. Then g restricted to components β and η is the identity, and the mapping of $(\pi_T)_{T \in \mathcal{C}^*}$ onto $(\pi_K)_{K \in \mathcal{K}^*}$ is defined through (8). This implies that

the performance state probabilities satisfy constraint (C1). Overall we have the composition of functions

$$\Theta_{\mathcal{C}} \xrightarrow{g} \Theta_{\mathcal{K}} \xrightarrow{f} \Phi$$

with $f \circ g$ denoting the prediction function of the CBLIM. Because g is onto by definition, the composition $f \circ g$ is one-to-one if and only if both functions f and g are one-to-one.

Proposition 7. *The following assertions are equivalent.*

1. *The function g is locally one-to-one at some point in $\Theta_{\mathcal{C}}$;*
2. *The function p is one-to-one;*
3. *The function g is one-to-one.*

Proof. The implication from 1. to 2. is shown by proving its contrapositive. Assume that p is not one-to-one. Then for some $K \in \mathcal{K}$ there are $T_1, T_2 \in \mathcal{C}$, $T_1 \neq T_2$, such $p(T_1) = p(T_2) = K$. We have to show that g is *not* locally one-to-one at all points in the parameter space. So, let $\theta_{\mathcal{C}} = (\beta, \eta, \dots, \pi_{T_1}, \dots, \pi_{T_2}, \dots)$ be an arbitrary point in $\Theta_{\mathcal{C}}$ and $\varepsilon > 0$. Because $P_{\mathcal{K}}(K)$ is nonzero, we can find a $\delta > 0$ such that $\theta'_{\mathcal{C}} = (\beta, \eta, \dots, \pi_{T_1} + \delta, \dots, \pi_{T_2} - \delta, \dots)$ lies in $\Theta_{\mathcal{C}}$ and within distance ε from $\theta_{\mathcal{C}}$. But then $g(\theta_{\mathcal{C}}) = g(\theta'_{\mathcal{C}})$ and g is not locally one-to-one at $\theta_{\mathcal{C}}$.

For proving the implication from 2. to 3. suppose p is one-to-one. Then for all $T \in \mathcal{C}^*$ with $p(T) = K \in \mathcal{K}^*$ we obtain $\pi_T = \pi_K$ by (8), and thus $(\pi_T)_{T \in \mathcal{C}^*} = (\pi_K)_{K \in \mathcal{K}^*}$ because p is onto and maps S into Q . This immediately provides that $g(\theta_{\mathcal{C}}) = g(\theta'_{\mathcal{C}})$ implies $\theta_{\mathcal{C}} = \theta'_{\mathcal{C}}$ for all $\theta_{\mathcal{C}}, \theta'_{\mathcal{C}} \in \Theta_{\mathcal{C}}$.

The remaining implication is obvious. □

Proposition 7 shows that g cannot be locally one-to-one at any point without p being one-to-one. The latter, however, provides that g is globally one-to-one. This leads to the situation that g is globally one-to-one if and only if it is locally one-to-one at some point, a very special situation indeed. The conclusions to be drawn from this result are immediate.

Corollary 1. *A CBLIM is identifiable if and only if the induced BLIM is identifiable and the problem function p is one-to-one.*

This means that a CBLIM cannot be identifiable whenever either p is not one-to-one, or the induced BLIM is not identifiable. Local identifiability of the BLIM can be checked by determining the rank of the Jacobian matrix of its prediction function. Stefanutti et al. (2012) provide an implementation of the necessary computations. This allows for testing local identifiability of a BLIM up to moderate size, which, in case it does not hold, leads to the conclusion that the BLIM is not identifiable. If the BLIM turns out to be locally identifiable and the problem function p is one-to-one, we may conclude that the CBLIM is locally identifiable, too. To be precise, we have the following result.

Proposition 8. *The CBLIM is locally identifiable at a point θ in $\Theta_{\mathcal{C}}$ if and only if the induced BLIM is locally identifiable at the point $g(\theta)$ in $\Theta_{\mathcal{K}}$ and the problem function p is one-to-one.*

Proof. First, notice that since f and g are analytic functions, it follows that $f \circ g$ is analytic, too. Thus, in particular, all of these functions are continuous.

Let $f \circ g$ be one-to-one if restricted to all points within a distance $\delta > 0$ from θ in $\Theta_{\mathcal{C}}$, and let $\mathcal{N}_{\theta}^{\delta}$ denote the set of these points. It is clear that g restricted to $\mathcal{N}_{\theta}^{\delta}$ is one-to-one, and thus by Prop. 7 the associated problem function p is one-to-one. Then f is one-to-one on $g(\mathcal{N}_{\theta}^{\delta})$. Since the inverse function g^{-1} on $g(\mathcal{N}_{\theta}^{\delta})$ exists and is continuous, we can select $\varepsilon > 0$ such that

$\mathcal{N}_{g(\theta)}^\epsilon \subseteq g(\mathcal{N}_\theta^\delta)$. This shows that f is one-to-one on $\mathcal{N}_{g(\theta)}^\epsilon$ and the BLIM is locally identifiable at $g(\theta)$.

Conversely, let f be one-to-one if restricted to all points within a distance $\epsilon > 0$ from $g(\theta)$ in $\Theta_{\mathcal{K}}$, i.e., if restricted to $\mathcal{N}_{g(\theta)}^\epsilon$, and let p be one-to-one. Then g is continuous and, by Prop. 7, one-to-one. This means that there is a $\delta > 0$ such that for all points $\theta_{\mathcal{C}}$ in $\Theta_{\mathcal{C}}$ within distance δ from θ the points $g(\theta_{\mathcal{C}})$ are within distance ϵ from $g(\theta)$, which implies that the function $f \circ g$ is one-to-one if restricted to all points within distance δ from θ . \square

The following is looking for necessary and sufficient conditions on a skill function μ that make the corresponding problem function one-to-one. In order to prepare the generalization suggested in the subsequent section, we need the notion of an atom of a competence structure \mathcal{C} . A subset $A \in \mathcal{C}$ is an *atom at* $s \in S$ if it is a minimal set in \mathcal{C} containing s . It is called an *atom* if it is an atom at s for some $s \in S$ (cf. Falmagne & Doignon, 2011). With this concept at hand we consider the following property. A skill function (Q, S, μ) respects the *witness condition* with respect to the competence structure \mathcal{C} whenever

for every atom $A \in \mathcal{C}$ there is some item $q \in Q$ such that $A \in \mu(q)$. (W-C)

In the so far considered case of $\mathcal{C} = 2^S$ the atoms are the singleton sets $\{s\}$ for all $s \in S$. In this situation the witness condition, which will be denoted by (W- 2^S), is easily translated into a property of the \mathbf{Q} -matrix corresponding to a conjunctive skill function: for every column (skill) j of \mathbf{Q} there is some row (item) i such that $\mathbf{Q}_{ij} = 1$, and $\mathbf{Q}_{ik} = 0$ for every other column $k \neq j$. Condition (W- 2^S) essentially says that for each skill there is at least one item which can be solved by exactly that skill. The following result shows that, with any arbitrary skill function, the witness condition (W-C) with respect to some competence structure \mathcal{C} is necessary for the problem function to be one-to-one on \mathcal{C} , and sufficient for conjunctive skill functions.

Proposition 9. *Let (Q, S, μ) be a skill function, p the corresponding problem function, and \mathcal{C} a competence structure on S . If p is one-to-one on \mathcal{C} then μ respects (W-C), and whenever μ is a conjunctive skill function respecting (W-C) then p is one-to-one on \mathcal{C} .*

Proof. To show necessity, suppose that μ does not respect (W-C). Then there is some atom $A \in \mathcal{C}$ such that $p(A) = \emptyset$, meaning that there are at least two competence states (namely \emptyset and A) that are mapped onto the same performance state \emptyset and thus p is not one-to-one on \mathcal{C} .

The proof of sufficiency for conjunctive skill functions proceeds by contradiction. Suppose p is not one-to-one on \mathcal{C} . Then there are distinct competence states C_1, C_2 in \mathcal{C} with $p(C_1) = p(C_2)$. Without loss of generality there is an atom $A \subseteq C_1$ in \mathcal{C} , which is not in C_2 . Now, assume that the conjunctive skill function μ respects (W-C). Then there is a $q \in Q$ such that $A \in \mu(q)$. But since $q \in p(C_2)$ there is a subset $T \subseteq C_2$ with $T \in \mu(q)$ and $T \neq A$, a contradiction. So μ cannot respect (W-C). \square

The subsequent result sheds light on the role of the witness condition from a structural point of view. For this we need the following notion: Given two partial orders (X, \leq_X) and (Y, \leq_Y) a bijective mapping $h: X \rightarrow Y$ is an *order-isomorphism* whenever for all $x, x' \in X$ we have $x \leq_X x'$ if and only if $h(x) \leq_Y h(x')$.

Proposition 10. *Let Q be a knowledge domain, S a set of skills, μ a conjunctive skill function, and \mathcal{C} a competence structure on S . Then μ respects (W-C) if and only if its induced problem function p is an order-isomorphism from \mathcal{C} to \mathcal{K} (with respect to \subseteq).*

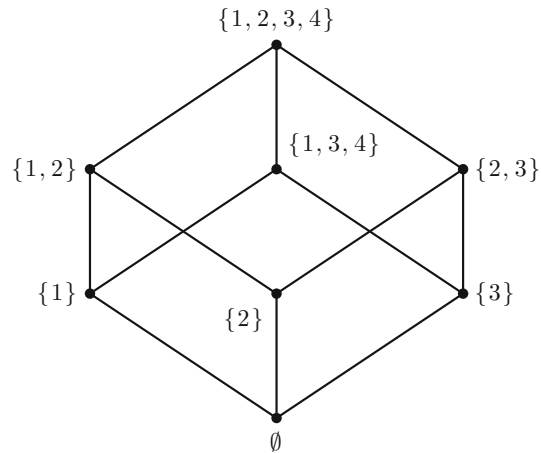


FIGURE 1.
Example.

Proof. Let μ satisfy (W-C). Then by definition $C_1 \subseteq C_2$ implies $p(C_1) \subseteq p(C_2)$ for all $C_1, C_2 \in \mathcal{C}$ and p is onto \mathcal{K} . Moreover, p is one-to-one on \mathcal{C} by Prop. 9. It remains to show that $p(C_1) \subseteq p(C_2)$ implies $C_1 \subseteq C_2$ for all $C_1, C_2 \in \mathcal{C}$. Assume that $p(C_1) \subseteq p(C_2)$ holds for $C_1, C_2 \in \mathcal{C}$, and let $s \in C_1$. Consider any atom A of \mathcal{C} with $s \in A$. Then by (W-C) there is an item $q \in \mathcal{Q}$ such that $\mu(q) = \{A\}$. By assumption $q \in p(C_2)$, which means that $A \subseteq C_2$ and thus $s \in C_2$.

Conversely, assume that p is an order-isomorphism from \mathcal{C} to \mathcal{K} . Let A be an atom of \mathcal{C} . Then obviously for any $q \in p(A)$ we have $\mu(q) = \{A\}$. \square

Example 3. Let the conjunctive skill function μ be defined on $\mathcal{Q} = \{1, 2, 3, 4\}$ and $S = \{a, b, c\}$ by $\mu(1) = \{\{a\}\}$, $\mu(2) = \{\{b\}\}$, $\mu(3) = \{\{c\}\}$, $\mu(4) = \{\{a, c\}\}$. It is easily seen that the problem function $p: \mathcal{C} \rightarrow 2^{\mathcal{Q}}$ with $\mathcal{C} = 2^S$ induced by μ is one-to-one. The delineated knowledge structure $p(\mathcal{C}) = \mathcal{K}$ is order-isomorphic to the powerset 2^S , which is obvious from its Hasse diagram depicted in Figure 1.

5.2. Toward Restoring Identifiability

Identifiability issues have been recognized in CDM from very early on (Tatsuoka, 1991). In case of $\mathcal{C} = 2^S$, Proposition 9 formulates the theoretical basis for a simple recipe, which has been suggested for restoring identifiability of the DINA model (see, e.g., Tatsuoka, 1990; DeCarlo, 2011). For a conjunctive skill function μ , suppose one has two items q and r and two skills a and b such that item q can be solved by skill a (i.e. $\mu(q) = \{\{a\}\}$) whereas item r requires both a and b ($\mu(r) = \{\{a, b\}\}$). Then (W- 2^S) requires the existence of a third item that can be solved by only skill b . At first sight, this restriction looks rather innocent. In those cases where the condition is violated, it seems to be just a matter of adding items as necessary for restoring the witness condition. This, however, is too optimistic. There might be empirical settings where these necessary items are difficult to find, or simply do not exist. Just think of a skill that can be applied only after another skill. Then it will be impossible to find an item that only requires the first, but not the second skill (see also DeCarlo, 2011). Tatsuoka (2009) raises still some more practical concerns. However, Corollary 1 and Proposition 8 make clear that even successfully applying this recipe will fail to restore identifiability of a CBLIM whenever its induced BLIM is

not identifiable. This is true, for instance, for the CBLIM of Example 3. The rank of the Jacobian matrix of its induced BLIM (see Figure 1) equals 13, while there are 15 parameters.

If the skill function μ does not respect (W- \mathcal{C}), then there are distinct competence states in \mathcal{C} delineating the same performance state, so that any assessment based on the latter remains ambiguous. Formally, with p denoting the corresponding problem function, this amounts to considering an equivalence relation \sim_p on the competence states. Defining

$$C_1 \sim_p C_2 \text{ if and only if } p(C_1) = p(C_2) \quad (15)$$

for all $C_1, C_2 \in \mathcal{C}$, the resulting set of equivalence classes \mathcal{C}/\sim_p is partially ordered by

$$[C_1]_p \sqsubseteq [C_2]_p \text{ if and only if } p(C_1) \subseteq p(C_2). \quad (16)$$

Obviously, the induced mapping $p^*: \mathcal{C}/\sim_p \rightarrow \mathcal{K}$ establishes a one-to-one correspondence. Notice that Tatsuoka (1996, 2002) suggested an analogous approach in the CDM context. The probabilistic framework on \mathcal{C}/\sim_p can simply be identified with the induced BLIM via the mapping p^* . This means that the probability distribution on the equivalence classes in \mathcal{C}/\sim_p is identical to that on the delineated knowledge structure \mathcal{K} , and can be estimated through well-established methods for the BLIM (Heller and Wickelmaier 2013).

Another way to arrive at a one-to-one correspondence between competence states and performance states is to restrict the power set on the set of skills S to only those subsets that are plausible to occur. Such an approach no longer assumes the skills to be independent, but introduces some kind of structure upon them. This idea was captured by the notion of a competence structure on the skills in KST (Korossy, 1997, 1999). A competence structure has already been introduced above as a collection of subsets of skills containing at least \emptyset and S . Similar ideas have been adopted in CDM only recently (de la Torre, Hong, & Deng, 2010). Notice that the results in Section 5.1 were formulated for arbitrary competence structures, and thus still apply to this general situation.

There are particularly nice results for the conjunctive CBLIM, or DINA model, respectively. In this special case the equivalence classes $[C]_p$ in \mathcal{C}/\sim are closed under intersection as a direct consequence of (3). The intersection of all the competence states that delineate a certain performance state K forms the minimal set of skills that are both necessary and sufficient for solving exactly the items in K . Thus, referring to this set as the unique representative of the equivalence class not only induces a one-to-one correspondence between performance and competence states, but also identifies the skills that need to be available for certain. Let \mathcal{C}_p denote the collection of all those minimal competence states, i.e., the set $\{\bigcap[T]_p : T \subseteq S\}$.

Example 4. Let (Q, S, μ) be a conjunctive skill function with $Q = \{1, 2, 3, 4\}$, $S = \{a, b, c, d\}$, and $\mu(1) = \{\{a\}\}$, $\mu(2) = \{\{a, b, c\}\}$, $\mu(3) = \{\{a, b, d\}\}$, $\mu(4) = \{\{a, c, d\}\}$.

The reader may easily verify that the following equalities hold true:

$$p(\{a\}) = p(\{a, b\}) = p(\{a, c\}) = p(\{a, d\}) = \{1\}.$$

The subset $\{a\}$ contains all the skills that are necessary and sufficient for solving exactly item 1. Any other subset contains skills that are either not sufficient (e.g., $\{b, c, d\}$), or not necessary (e.g., $\{a, b\}$) for solving item 1. Moreover, $\{a\}$ is the unique minimal competence state corresponding to $\{1\}$. The knowledge structure delineated by μ is given by

$$\mathcal{K} = p(2^S) = \{\emptyset, \{1\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, Q\},$$

and the corresponding collection of minimal competence states by

$$\mathcal{C}_p = \{\emptyset, \{a\}, \{a, b, c\}, \{a, b, d\}, \{a, c, d\}, S\}.$$

Notice that $(W-\mathcal{C}_p)$ is respected, because the atoms $\{a\}, \{a, b, c\}, \{a, b, d\}, \{a, c, d\}$ in \mathcal{C}_p are contained in $\mu(1), \mu(2), \mu(3), \mu(4)$, respectively. Obviously, the two collections are in a one-to-one correspondence.

The following provides a structural characterization of the competence structure \mathcal{C}_p and an answer to the question which type of structure is imposed on the skills.

Proposition 11. *Let p be the problem function corresponding to a conjunctive skill function (Q, S, μ) . Then the collection \mathcal{C}_p is closed under union and thus forms a competence space.*

Proof. Let $C, C' \in \mathcal{C}_p$ be any two minimal competence states and let $M \subset C \cup C'$. Then at least one of the two inequalities $C \setminus M \neq \emptyset$ or $C' \setminus M \neq \emptyset$ holds true. Assuming without loss of generality the first of them, since C is minimal for $p(C)$, there must be $q \in p(C)$ which is not in $p(M)$. But, for the fact that p is order preserving, $q \in p(C)$ implies $q \in p(C \cup C')$. Hence $p(M) \neq p(C \cup C')$. This shows that $C \cup C'$ must be the minimal element delineating $p(C \cup C')$. \square

Notice that the collection \mathcal{C}_p needs not be closed under intersection. A counterexample is provided by the competence space in Example 4: both $\{a, b, c\}$ and $\{a, b, d\}$ are in \mathcal{C}_p , but their intersection is not.

If the competence space \mathcal{C}_p is distinct from the powerset 2^S then there are dependencies between the skills. These dependencies may reflect, for example, that the skills are acquired in a certain order. Formally, they are described by relating subsets of skills. There is a one-to-one correspondence between competence spaces \mathcal{C}_p and binary relations \mathcal{P} on $2^S \setminus \{\emptyset\}$ that are transitive and extend the inverse set inclusion \supseteq (Doignon and Falmagne, 1999 Theorems 5.5 and 5.7). We have $T_1 \mathcal{P} T_2$ for some subsets $T_1, T_2 \subseteq S$ if and only if the following implication holds: If the skills in T_1 are not available, then the skills in T_2 are not available, too. This order generalizes the binary prerequisite relation on the set S that de la Torre et al. (2010) assumed to capture a hierarchical structure on the skills.

6. Empirical Example

This section illustrates the application of the above derived results. The data considered below consists of the responses that 536 middle school students gave to 15 fraction subtraction items, and forms a subset of the original data described by Tatsuoka (1990). This data set has already been analyzed by de la Torre and Douglas (2008) based on the DINA and MS-DINA models, assuming seven skills to be relevant for solving the items. The skills were defined as follows: (a) Performing basic fraction subtraction operation, (b) simplifying/reducing, (c) separating whole number from fraction, (d) borrowing one from whole number to fraction, (e) converting whole number to fraction, (f) converting mixed number to fraction, and (g) column borrowing in subtraction. The skill assignment was defined by the \mathbf{Q} -matrices \mathbf{Q}_1 and \mathbf{Q}_2 given in Table 3. The two \mathbf{Q} -matrices specify alternative strategies that could be used for solving each item. For example, a correct response to Item 12 can be obtained in two ways, one requiring skills a, c , and d , and the other requiring skills a, b , and f . Only a single strategy is assumed for Item 8, which requires skills a , and b . Notice that \mathbf{Q}_1 corresponds to what Tatsuoka (1990) called Method B, while \mathbf{Q}_2 corresponds

TABLE 3.
 \mathbf{Q} -matrices \mathbf{Q}_1 and \mathbf{Q}_2 as well as conjunctive skill function μ_1 , and skill function $\mu_{1,2}$.

Item	\mathbf{Q}_1					\mathbf{Q}_2					Conjunctive	
	a	b	c	d	e	a	b	e	f	g	Skill function	Skill multimap
1	1	0	0	0	0	1	0	0	0	0	$\mu_1(1) = \{\{a\}\}$	$\mu_{1,2}(1) = \{\{a\}\}$
2	1	1	1	1	0	1	0	0	1	0	$\mu_1(2) = \{\{a, b, c, d\}\}$	$\mu_{1,2}(2) = \{\{a, b, c, d\}, \{a, f\}\}$
3	1	0	0	0	0	1	0	0	0	0	$\mu_1(3) = \{\{a\}\}$	$\mu_{1,2}(3) = \{\{a\}\}$
4	1	1	1	1	1	1	0	1	1	0	$\mu_1(4) = \{\{a, b, c, d, e\}\}$	$\mu_{1,2}(4) = \{\{a, b, c, d, e\}, \{a, e, f\}\}$
5	0	0	1	0	0	0	1	1	1	1	$\mu_1(5) = \{\{c\}\}$	$\mu_{1,2}(5) = \{\{c\}, \{b, e, f, g\}\}$
6	1	1	1	1	0	1	1	0	1	0	$\mu_1(6) = \{\{a, b, c, d\}\}$	$\mu_{1,2}(6) = \{\{a, b, c, d\}, \{a, b, f\}\}$
7	1	1	1	1	0	1	1	0	1	0	$\mu_1(7) = \{\{a, b, c, d\}\}$	$\mu_{1,2}(7) = \{\{a, b, c, d\}, \{a, b, f\}\}$
8	1	1	0	0	0	1	1	0	0	0	$\mu_1(8) = \{\{a, b\}\}$	$\mu_{1,2}(8) = \{\{a, b\}\}$
9	1	0	1	0	0	1	0	0	1	0	$\mu_1(9) = \{\{a, c\}\}$	$\mu_{1,2}(9) = \{\{a, c\}, \{a, f\}\}$
10	1	0	1	1	1	1	1	1	0	0	$\mu_1(10) = \{\{a, c, d, e\}\}$	$\mu_{1,2}(10) = \{\{a, c, d, e\}, \{a, b, e\}\}$
11	1	0	1	0	0	1	1	0	1	0	$\mu_1(11) = \{\{a, c, \}\}$	$\mu_{1,2}(11) = \{\{a, c, \}, \{a, b, f\}\}$
12	1	0	1	1	0	1	1	0	1	0	$\mu_1(12) = \{\{a, c, d\}\}$	$\mu_{1,2}(12) = \{\{a, c, d\}, \{a, b, f\}\}$
13	1	1	1	1	0	1	1	0	1	1	$\mu_1(13) = \{\{a, b, c, d\}\}$	$\mu_{1,2}(13) = \{\{a, b, c, d\}, \{a, b, f, g\}\}$
14	1	1	1	1	1	1	1	1	1	1	$\mu_1(14) = \{\{a, b, c, d, e\}\}$	$\mu_{1,2}(14) = \{\{a, b, c, d, e\}, \{a, b, e, f, g\}\}$
15	1	1	1	1	0	1	1	0	1	1	$\mu_1(15) = \{\{a, b, c, d\}\}$	$\mu_{1,2}(15) = \{\{a, b, c, d\}, \{a, b, f, g\}\}$

See text for definition of skills.

to her Method A, and the skills used as well as their assignment to the items differs slightly from the original paper.

Including a higher-order latent trait (which relates mastery of the skills to a unidimensional latent trait according to de la Torre & Douglas, 2004) into the model, de la Torre and Douglas (2008) estimated the DINA on \mathbf{Q}_1 and the MS-DINA on the collection $\mathcal{M} = \{\mathbf{Q}_1, \mathbf{Q}_2\}$. In the present work, the BLIM is estimated on the performance structures delineated by the conjunctive skill function μ_1 on the skill set $S = \{a, b, c, d, e\}$ corresponding to \mathbf{Q}_1 and by the skill function $\mu_{1,2}$ on $S = \{a, b, c, d, e, f, g\}$ corresponding to $\mathcal{M} = \{\mathbf{Q}_1, \mathbf{Q}_2\}$ (see Table 3). When comparing these results with those found by de la Torre and Douglas (2008) the reader should bear in mind that there are substantial methodological differences. While de la Torre and Douglas (2008) used a higher-order latent trait specification of the models and MCMC parameter estimation, the reported analysis employed an EM algorithm as implemented in the package ‘pks’ (Heller & Wickelmaier, 2013) of the R environment for statistical computing (R Core Team, 2013).

The DINA model and the CBLIM defined through the conjunctive skill function μ_1 are considered first. On the one hand, it is easily seen that (W-2^S) is not respected (for none of the five skills in \mathbf{Q}_1 there is one item that can be solved by exactly that skill), and thus the correspondence between competence and performance states is many-to-one. Indeed, the 32 competence states delineate only 10 different performance states (see the Hasse diagram of \mathcal{K} in Figure 2). The four competence states $\{a, b\}$, $\{a, b, d\}$, $\{a, b, e\}$, and $\{a, b, d, e\}$, for example, are all mapped onto the performance state $\{1, 3, 8\}$. On the other hand, the rank of the Jacobian matrix of the prediction function of the induced BLIM equals the number of free parameters $2 \cdot 15 + 10 - 1 = 39$. This implies that the BLIM is locally identifiable, whereas the respective CBLIM and DINA model are *not* identifiable. While the probabilities of the performance states are uniquely determined, there is no indication at all of how to divide these probabilities among the members of the corresponding equivalence classes. Given the local identifiability of the BLIM, identifiability of the CBLIM and DINA model can be restored by considering the collection \mathcal{C}_p (see Figure 2), and by assigning to each of the unique minimal competence states the probability of the performance state it

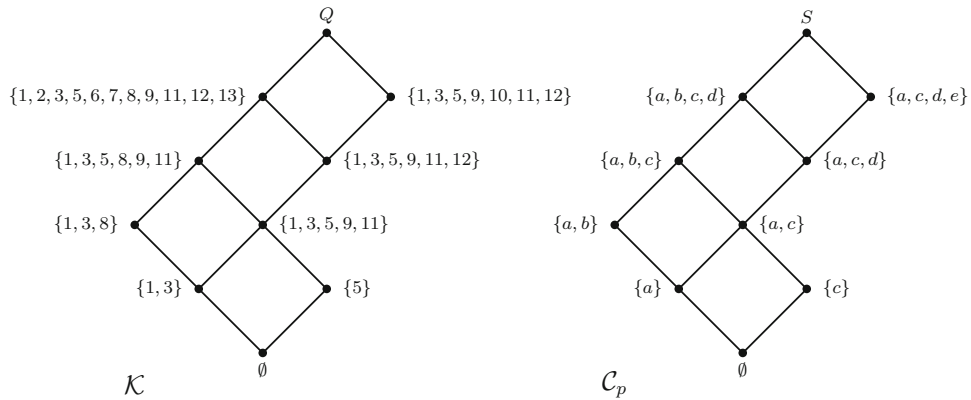


FIGURE 2.

Hasse diagrams of the delineated knowledge structure \mathcal{K} and the collection \mathcal{C}_p of minimal competence states for the conjunctive skill function μ_1 corresponding to the \mathbf{Q} -matrix \mathbf{Q}_1 .

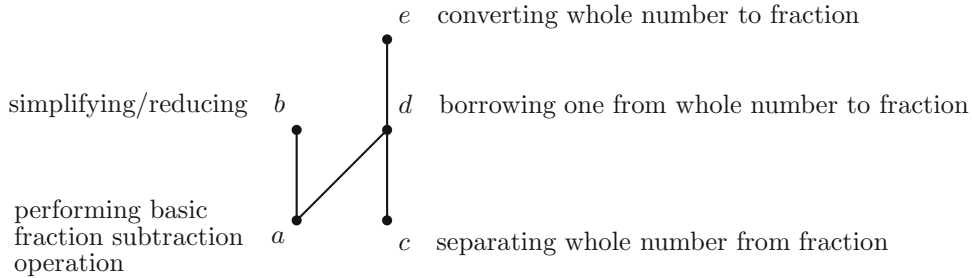


FIGURE 3.

Hasse diagram of the prerequisite relation equivalent to the competence structure \mathcal{C}_p from Figure 2.

delineates. In this particular case the skill structure inherent in \mathcal{C}_p can be represented by a binary prerequisite relation on S (Doignon & Falgagne, 1999, Theorem 1.49), which is illustrated in Figure 3. Notice that in general the skill structure is captured by a binary relation on $2^S \setminus \{\emptyset\}$ rather than on S (see end of Section 5.2). The depicted relation shows that, for example, ‘performing basic fraction subtraction operation’ (skill a) and ‘separating whole number from fraction’ (skill c) are prerequisites to ‘borrowing one from whole number to fraction’ (skill d). This actually reflects the fact that all items requiring skill d also require skills a and c (we note in passing that the dependence of d on c was already mentioned in Klein, Birenbaum, Standiford, & Tatsuoka, 1990). The above mentioned duality of skills might lead one to suspect that an individual having available skill d also has available skills a and c . This, however, does not necessarily follow and needs further scrutiny.

The equivalence of the two models is also mirrored at a more practical level, when it comes to parameter estimation. The lucky guess and careless error probabilities estimated for the CBLIM resemble the guessing and slip probabilities reported by de la Torre and Douglas (2008) for the DINA model. The discrepancies are negligible (>0.01 in one out of 30 cases). Based on the estimated parameters, expected proportions of correct responses to each item and expected log-odds ratios between the items were simulated for a large data set (100,000 response patterns). The results are very close to those obtained by de la Torre and Douglas with an analogous procedure.

The mean absolute difference is smaller than 0.01 for the expected proportions of correct responses and equals 0.03 for the log-odds ratios.

Considering the MS-DINA model and the CBLIM based on $\mathcal{M} = \{\mathbf{Q}_1, \mathbf{Q}_2\}$, the skill function $\mu_{1,2}$ does not respect $(W-2^S)$, and a total of 128 competence states delineates no more than 23 performance states. As the rank of the Jacobian matrix of the prediction function of the induced BLIM equals the number of free parameters $2 \cdot 15 + 23 - 1 = 52$, this again implies that the BLIM is locally identifiable, whereas the respective CBLIM and MS-DINA model are *not* identifiable. Restoring identifiability along the lines outlined above, however, is not possible. The five competence states $\{a, b, c, d\}$, $\{a, b, c, d, g\}$, $\{a, b, c, d, f\}$, $\{a, b, c, f, g\}$, and $\{a, b, c, d, f, g\}$, for example, all delineate the performance state $\{1, 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 15\}$, but this equivalence class is *not* closed under intersection (e.g., $\{a, b, c, d\} \cap \{a, b, c, f, g\}$ is not in the equivalence class). The estimated response error probabilities for the CBLIM largely resemble those reported by [de la Torre and Douglas \(2008\)](#) for the MS-DINA model. Discrepancies >0.02 occur in four out of 30 cases and may be attributed to the methodological differences mentioned above. In a simulation as described above for the DINA model the mean absolute difference is smaller than 0.01 for the expected proportions of correct responses and is 0.08 for the log-odds ratios.

7. Conclusions

The present paper provides a first systematic comparison of theories developed within the frameworks offered by KST and CDM. It pinpoints to the correspondences between the two approaches, ranging from the very basic concepts to even specific models. These correspondences have so far only been alluded to. One of the reasons for this might be that KST originally does not refer to skills, or attributes ([Doignon & Falmagne, Doignon and Falmagne 1985](#)). However, once this is remedied, the respective counterparts of notions and theories in each of the approaches are easily identified.

The correspondences are spelled out for the MS-DINA model on part of the CDM, and the CBLIM, which is a skill-based extension of the standard probabilistic model (BLIM) within KST. It is shown that these models are equivalent. Moreover, the above results also establish equivalences between the DINA and DINO model on the one hand, and the conjunctive and disjunctive CBLIM on the other hand. Working out the exact correspondences between the two theories is not just a scientific exercise, but creates synergy effects, from which both camps can benefit. Progress can emerge from integrating the different perspectives that the two research strands bring in. While CDM adopts a more computational view, often but not always based on numerical vector representations, KST entails a more structural perspective based on set-theoretical representations. The latter perspective has also been taken up in parts of CDM, because vector representations seem to be inappropriate for treating identifiability issues (e.g., [Tatsuoka, 2002](#)).

The previous sections demonstrate that a fairly complete picture on the identifiability of the considered class of models emerges from combining recent results on probabilistic knowledge structures ([Heller, 2015](#); [Spoto, Stefanutti, & Vidotto, 2012](#); [Stefanutti et al., 2012](#)) with approaches taken in CDM (e.g., [Tatsuoka, 1996, 2002](#)). It is shown that there are two quite independent sources of non-identifiability: the structural aspects of the delineated performance structure, and the relation between performance and competence states (see [Corollary 1](#) and [Proposition 8](#)). Although, from the CDM perspective, at first glance it may appear that explicitly referring to the performance level introduces an additional layer which complicates matters, the above makes clear that the performance states carry important information about the competence states (as captured by, e.g., [\(15\)](#) and [\(16\)](#)). In a way, this has already been acknowledged in the CDM context, when addressing the many-to-one correspondence between competence and performance states as a source of non-identifiability ([Tatsuoka, 1996, 2002](#)). [Section 5.2](#) shows that

besides extending the set of items in order to respect the witness condition ($W-2^S$), which has already been discussed as a potential remedy (e.g., Tatsuoka, 1990; DeCarlo, 2011), its newly introduced generalization ($W-C$) allows for putting restrictions on the set of possible competence states. Restricting \mathcal{C} has been considered in KST (Korossy, 1997, 1999) and CDM (de la Torre et al., 2010) before, but not as a means for restoring identifiability. Due to its quasi-automatic nature, such a restriction to minimal competence states that are in one-to-one correspondence to the performance states always exists for the DINA model/conjunctive CBLIM. Whether the resulting structure on the skills is psychologically plausible, or merely reflects properties of the item set used, however, remains to be scrutinized.

Compared to the above discussed models, the LCDM (Henson et al. 2009) and the GDINA model (de la Torre, 2011) provide more flexibility in explaining the observed behavior. These models allow for quantifying the item-specific effects of single skills and skill combinations within the framework of log-linear models. By this they can generate conditional distributions of the observed response given the available skills that are based on a more elaborate classification beyond the dichotomy ‘all required skills present’ vs. ‘at least one of the required skills missing.’ In the DINA, DINO, and MS-DINA models as well as the CBLIM the conditional probability of solving (not solving) item q within the set T of available skills is $1 - \beta$ (β , respectively) whenever T contains a competency for solving q , and η_q ($1 - \eta_q$, respectively) otherwise. This latter probability, however, may change, if some but not all of the skills sufficient for solving q are available. Tatsuoka et al. (2013) outline such a generalization in the CDM context. Let us exemplify how to handle this within the above presented approach by considering the \mathbf{Q} -matrix \mathbf{Q}_1 of Table 2 on the domain $\mathcal{Q} = \{p, q, r, s\}$. Here, the subset of skills $\{b, c\}$ is sufficient for solving item q , and the subsets \emptyset and $\{b\}$ both delineate the empty performance state \emptyset (i.e., function g is not one-to-one). Now, assume that the probability of solving item q departs from the guessing level if skill b is available. Then this can be accounted for by extending the domain to $\mathcal{Q}^e = \{p, q_1, q_2, r, s\}$ by splitting item q into two virtual items q_1 and q_2 , and by defining the \mathbf{Q} -matrix (or skill function, respectively) as in Table 4. For the BLIM on \mathcal{Q}^e it is then reasonable to introduce the parameter restriction $\eta_{q_1} = \eta_{q_2}$. In contrast to the problem function corresponding to \mathbf{Q}_1 , the problem function induced by the extended \mathbf{Q} -matrix \mathbf{Q}_1^e , and thus the function $g_e: \Theta_{\mathcal{C}^e} \rightarrow \Theta_{\mathcal{K}^e}$ mapping the parameter space on the extended competence structure into that of the extended performance structure is one-to-one. Concerning the original domain \mathcal{Q} the item q is considered to be solved whenever q_1 or q_2 is solved. Then, by the above assumptions, q is solved with probability $1 - \beta_{q_1}$ if both skills b and c are available, with probability $1 - \beta_{q_2}$ if b but not c is available, and with probability $\eta_{q_1} = \eta_{q_2}$ if skill b is not available. For deciding on the identifiability of the model, first consider the function f_e mapping the extended parameter space $\Theta_{\mathcal{K}^e}$ into the distributions of the response patterns on the extended domain \mathcal{Q}^e . Notice that, by merging the virtual items q_1 and q_2 into q , even if f_e is one-to-one, this need not be the case for the mapping f into the distributions of the response patterns on the original domain \mathcal{Q} . In the present example, however, already f_e is not one-to-one (rank of the Jacobian matrix equals 14, with a total of 16 parameters). This means that the generalized model is not identifiable.

This again shows that exclusively focussing on the one-to-one correspondence between performance and competence states (represented by the function g) neglects the source of non-identifiability that is related to the function f . The presented theoretical results imply that even if the just discussed recipes and generalizations establish a one-to-one function g , they fail to restore identifiability if the function f is not one-to-one. A CBLIM cannot be identifiable whenever its induced BLIM is not identifiable.

The just described situation is expected to be essentially the same, when considering even more general models that may be formulated within the framework provided by the LCDM or the GDINA. Representing their greater flexibility by introducing virtual items as outlined above may let the correspondence between performance and competence states to be one-to-one. At the

TABLE 4.
Extended \mathbf{Q} -matrix \mathbf{Q}_1^e (compare to \mathbf{Q}_1 in Table 2).

Item	\mathbf{Q}_1^e		
	a	b	c
p	1	0	0
q_1	0	1	1
q_2	0	1	0
r	0	0	1
s	1	1	0

same time, however, the increased complexity of the induced BLIM will contribute to its non-identifiability. This makes it difficult to deduce general statements, but the established link to the CBLIM offers a way to test the identifiability of specific models. Spoto et al. (2012) and Heller (2015) have characterized some of the properties of the delineated performance structure that cause the BLIM defined on it to be non-identifiable, while Stefanutti et al. (2012) implemented a procedure for assessing its local identifiability.

Clarifying the link between CDM and KST opens up a whole research agenda. Notions developed in one of the areas might be exploited in the other area, too. In KST, for example, the upwards directed line sequences in a Hasse diagram as in Figure 2 represent learning paths from the naïve state \emptyset to the state of full mastery (corresponding to the set Q or S , respectively). Of particular interest in this context is a certain class of performance structures, the so-called *well-graded knowledge spaces* or *learning spaces* in KST terminology (Falmagne & Doignon, 2011). These structures have demonstrated their usefulness for concretely assessing the knowledge state of an individual and for personalizing learning. Well-graded knowledge spaces emerge from pedagogically sound assumptions on the learning process (Cosyn & Uzun, 2009):

1. If a person with knowledge state K is ready to learn an item, then another person with knowledge state $L \supseteq K$ either has learned that item, or is also ready to learn it.
2. Learning proceeds in a stepwise, or element-by-element manner (as in the right, but not in the left Hasse diagram in Figure 2).

Moreover, stochastic Markov models have been developed to describe the learning process moving along these paths (Falmagne, 1994). These ideas might be useful for putting restrictions on CDMs, which even may have the side-effect of rendering them identifiable.

On the other hand, there are models in CDM that, apparently, do not have any correspondence in KST. The already mentioned LCDM (Henson et al. 2009) and GDINA (de la Torre, 2011) provide important examples. The above outlined way of how to handle the flexibility of these models by appropriately adapting the considered CBLIM is nothing more than a first step toward a comprehensive account. Developing a corresponding class of models within KST deserves future attention. Moreover, in CDM there is a wealth of statistical methods for parameter estimation and model selection (like, e.g., MCMC algorithms) that may prove useful in KST. Thus, further trying to bridge the gap between CDM and KST seems to be a promising enterprise. This article provides first evidence of the benefits that may arise. Hopefully, it can contribute to facilitating the communication between the two camps, leading to more rapid advances in both research areas.

References

Albert, D., Schrepp, M., & Held, T. (1994). Construction of knowledge spaces for problem solving in chess. In G. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 123–135). New York: Springer-Verlag.

- Bamber, D., & van Santen, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, 29, 443–473.
- Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, 44, 20–40.
- Bolt, D. (2007). The present and future of IRT-based cognitive diagnostic models (ICDMs) and related methods. *Journal of Educational Measurement*, 44(4), 377–383.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Cosyn, E., & Uzun, H. (2009). Note on two sufficient axioms for a well-graded knowledge space. *Journal of Mathematical Psychology*, 53(1), 40–42.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115–130.
- de la Torre, J. (2011). The generalized dina model framework. *Psychometrika*, 76, 179–199.
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333–353.
- de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, 73(4), 595–624.
- DeCarlo, L. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, 35, 8–26.
- de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the dina model. *Journal of Educational Measurement*, 47(2), 227–249.
- DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overview: Irt-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, 44(4), 285–291.
- Doignon, J.-P. (1994). Knowledge spaces and skill assignments. In G. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 111–121). New York: Springer-Verlag.
- Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23, 175–196.
- Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. Berlin, Heidelberg, New York: Springer-Verlag.
- Düntsch, I., & Gediga, G. (1995). Skills and knowledge structures. *British Journal of Mathematical and Statistical Psychology*, 48, 9–27.
- Falmagne, J.-C. (1994). Finite markov learning models for knowledge structures. In G. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 75–89). New York: Springer.
- Falmagne, J.-C., & Doignon, J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, 41, 1–23.
- Falmagne, J.-C., & Doignon, J.-P. (1988b). A markovian procedure for assessing the states of a system. *Journal of Mathematical Psychology*, 32, 232–258.
- Falmagne, J.-C., & Doignon, J.-P. (2011). *Learning spaces: Interdisciplinary applied mathematics*. Berlin, Heidelberg: Springer.
- Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johanessen, L. (1990). Introduction to knowledge spaces: How to build, test and search them. *Psychological Review*, 97, 204–224.
- Gediga, G., & Düntsch, I. (2002). Skill set analysis in knowledge structures. *British Journal of Mathematical and Statistical Psychology*, 55, 361–384.
- Haertel, E. H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement*, 8, 333–346.
- Haertel, E. H. (1989). Using restricted latent class models to map skill structure of achievement items. *Journal of Educational Measurement*, 26, 301–321.
- Haertel, E. H. (1990). Continuous and discrete latent structure models of item response data. *Psychometrika*, 55, 477–494.
- Heller, J. (2015). Identifiability in probabilistic knowledge structures. Manuscript under revision.
- Heller, J., & Repitsch, C. (2008). Distributed skill functions and the meshing of knowledge structures. *Journal of Mathematical Psychology*, 52, 147–157.
- Heller, J., & Repitsch, C. (2012). Exploiting prior information in stochastic knowledge assessment. *Methodology*, 8, 12–22.
- Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, 42, 49–56.
- 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.
- 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.
- Klein, M., Birenbaum, M., Standiford, S., & Tatsuoka, K. (1990). *Logical error analysis and construction of tests to diagnose student "bugs" in addition and subtraction of fractions*. Urbana, IL: Computer-based Education Research Laboratory. (Research Report NO. CERL-RR-81-6).
- Korossy, K. (1997). Extending the theory of knowledge spaces: A competence-performance approach. *Zeitschrift für Psychologie*, 205, 53–82.
- Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, applications* (pp. 103–132). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187–212.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379–416.

- R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. <http://www.R-project.org/>
- Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement, 44*, 293–311.
- Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.
- Spoto, A., Stefanutti, L., & Vidotto, G. (2012). On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures. *Journal of Mathematical Psychology, 56*, 248–255.
- Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E. (2012). Assessing local identifiability of probabilistic knowledge structures. *Behavior Research Methods, 44*, 1197–1211.
- Tatsuoka, C. (1996). *Sequential classification on partially ordered sets*. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.
- Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society Series C (Applied Statistics), 51*, 337–350.
- Tatsuoka, C. (2009). Diagnostic models as partially ordered sets. *Measurement, 7*, 49–53.
- Tatsuoka, C., Varadi, F., & Jaeger, J. (2013). Latent partially ordered classification models and normal mixtures. *Journal of Educational and Behavioral Statistics, 38*, 267–294.
- Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453–488). Hillsdale: Lawrence Erlbaum Associates.
- Tatsuoka, K. (1991). *Boolean algebra applied to determination of the universal set of misconception states*. Princeton, NJ: Educational Testing Services. (ONR-Technical Report No. RR-91-44).
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*, 287–305.
- von Davier, M. (2005). *A general diagnostic model applied to language testing data*. Princeton, NJ: Educational Testing Service. (Research Report No. RR-05-16).

Manuscript Received: 15 OCT 2012

Published Online Date: 3 APR 2015