A class of k -modes algorithms for extracting knowledge structures from data

Debora de Chiusole¹ · Luca Stefanutti¹ · Andrea Spoto²

Published online: 29 August 2016
© Psychonomic Society, Inc. 2016

Abstract One of the most crucial issues in knowledge space theory is the construction of the so-called knowledge structures. In the present paper, a new data-driven procedure for large data sets is described, which overcomes some of the drawbacks of the already existing methods. The procedure, called k -states, is an incremental extension of the k -modes algorithm, which generates a sequence of locally optimal knowledge structures of increasing size, among which a “best” model is selected. The performance of k -states is compared to other two procedures in both a simulation study and an empirical application. In the former, k -states displays a better accuracy in reconstructing knowledge structures; in the latter, the structure extracted by k -states obtained a better fit.

Keywords Knowledge structures · k -modes · Data-driven procedures

Introduction

Knowledge space theory (KST; Doignon & Falmagne, 1985, 1999; Falmagne & Doignon, 2011), is a theory

developed in the field of mathematical psychology and fruitfully applied in adaptive knowledge assessment. In this theory, individual knowledge is represented by the so-called *knowledge state*, which is the subset K of all problems in a given domain Q that the individual is capable of solving. A *knowledge structure* is a pair $(Q, \mathcal{K})$ where $\mathcal{K}$ is a collection of knowledge states containing at least $\emptyset$ and Q itself. Some particular kinds of structures are defined on the basis of the specific closure they satisfy: whenever a knowledge structure is closed under union (i.e., every union of states is a state in the structure), it is named a *knowledge space*; a structure closed under intersection (i.e., every intersection of states is a state in the structure) is a *closure space*; finally, a structure closed under both union and intersection is a *quasi ordinal space*.

This article explores one of the more critical issues in KST, that is the construction of the so-called knowledge structures. This issue has been deeply explored within KST, in which three main categories of methods exist. The first one refers to the *query to experts* (Dowling, 1993; Kambouri et al., 1994; Koppen, 1993; Koppen & Doignon, 1990; Müller, 1989; Schrepp & Held, 1995). The second method is based on *skill map* construction (Heller et al., 2013; Albert & Lukas, 1999; Doignon, 1994; Lukas & Albert, 1993). The third category is that of *data-driven* approaches (Falmagne et al., 2013; Robusto & Stefanutti, 2014; Sargin & Ünlü, 2009; Schrepp, 1999a, b, 2003; Spoto et al., 2015; Villano, 1991).

Data-driven methods can be further classified into two categories. By imposing specific properties to the knowledge structure underlying the data, methods of the former category are capable of inferring knowledge states that are never observed in a data set (Sargin & Ünlü, 2009; Schrepp, 1999b, 2003; Spoto et al., 2015). On the contrary, methods of the second category do not impose any restrictions to the

✉ Andrea Spoto
andrea.spoto@unipd.it

Debora de Chiusole
debora.dechiusole@gmail.com

Luca Stefanutti
luca.stefanutti@unipd.it

¹ FISPPA Department, University of Padua, Padova, Italy

² Department of General Psychology, University of Padua, Via Venezia 8, 35131 Padova, PD, Italy

underlying knowledge structure, but they cannot infer the existence of states that have never been observed (Robusto & Stefanutti, 2014; Schrepp, 1999a).

In the present paper, a procedure for extracting a knowledge structure out a set of observed data is described. The procedure aims at constructing a knowledge structure by neither imposing restrictions on it, nor assuming that only observed patterns could be states. It is an incremental extension of the k -modes algorithm (Chaturvedi et al., 2001; Huang & Ng, 1999) to knowledge structure extraction.

The paper is organized as follows. After presenting in some more details the KST data-driven methodologies (“Data-driven methodologies in KST”), the k -modes algorithm is introduced (“The k -modes algorithm”). Then, an adaptation of k -modes to the KST framework is presented (“A k -modes approach to knowledge structure extraction”) along with its incremental extension (“Incremental extensions of k -modes”). This last version of the algorithm, called k -states, is then tested through a simulation study (“Simulation study”) and an empirical application (“Empirical application”) in which its performance is compared with those of other two KST data-driven methodologies developed by Schrepp (1999a, 2003). The paper concludes with a discussion (“Discussion”).

Data-driven methodologies in KST

It has to be stressed that the data-driven methods are aimed at building a knowledge structure through an empirical approach, without any prior theoretical assumptions about either the relations among items, or the skills needed to solve them. This aspect clearly distinguishes them from both the expert query and the skill-map approaches.

All procedures described in this section apply to a data set consisting of a collection of N response patterns, each of which is represented by the subset $R \subseteq Q$ of those items receiving a correct response.

All the data-driven methods proposed so far refer to two main categories (Falmagne et al., 2013): (a) the Boolean analysis of questionnaires methods, aimed at building an implication relation among the items of a questionnaire (Schrepp, 1999b; Sargin & Ünlü, 2009); and (b) the methods to derive structures directly from data (e.g.; Schrepp, 1999a, b; Desmarais et al., 1995). Examples of the former class of methods are the Item Tree Analysis (ITA; Ünlü & Albert, 2004; Van Leeuwe, 1974) and the Inductive Item Tree Analysis (IITA; Schrepp, 2003, 2002), while examples of the latter can be found in Schrepp (1999a, b). Furthermore, some “hybrid” procedures have been proposed that put together a data-driven approach and either skill maps (D-SMEP Spoto et al., 2015) or query to experts procedure (Cosyn & Thiéry, 2000). All methodologies follow a

three-step procedure: (a) constructing a set of knowledge structures; (b) testing the models according to a set of fitting criteria; and (c) selecting the best fitting model as the representation of the latent structure.

Since the knowledge states are extracted on the basis of the empirical evidence, existing data-driven procedures operate by either: (1) simply assuming that all knowledge states are observable (Robusto & Stefanutti, 2014; Schrepp, 1999a, b) or (2) imposing specific properties to the structure, that allow to infer the unobserved states. For example ITA and IITA are conceived to build quasi ordinal spaces, whereas D-SMEP builds closure spaces, knowledge spaces or quasi ordinal spaces.

Both ways described in 1 and 2 have drawbacks: on the one hand, the assumption that states must have a positive observed frequency could be false in finite samples. For instance, the observed response pattern could differ from the knowledge state underlying it because of careless errors or lucky guesses to some of the items. On the other hand, there might be empirical situations in which the assumptions on the properties of the extracted knowledge structure are too strict or false. For this reason, the procedure itself could lack generality.

The k -modes algorithm

The k -modes algorithm represents an extension of the k -means (Hartigan & Wong, 1979) paradigm to categorical data. Such extension is based on three fundamental characterizations: (a) k -modes uses a simple matching dissimilarity measure; (b) it refers to modes statistics instead of means to center clusters; and (c) it updates modes on the basis of observed frequencies. In the next paragraphs, an overview of these three crucial issues is provided.

Consider two categorical objects represented by the two m -dimensional vectors X and Y . A dissimilarity measure for X and Y can be obtained as the number of mismatching elements. The higher the number of mismatches, the less similar X and Y . The discrepancy measure d is formally expressed as:

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j)$$

with

$$\delta(x_j, y_j) = \begin{cases} 0 & \text{if } x_j = y_j \\ 1 & \text{if } x_j \neq y_j. \end{cases}$$

One of the most interesting aspects of k -modes, that turns to be very useful in the proposed application, is the definition of the mode of a set $\mathcal{X}$ of objects X . In fact the mode is defined as the object $Q = (q_1, q_2, \dots, q_m)$ that minimizes

the distance D to $\mathcal{X}$. Formally the distance D is defined as follows:

$$D(\mathcal{X}, \mathcal{Q}) = \sum_{i=1}^n d(X_i, \mathcal{Q}),$$

where n is the number of objects of $\mathcal{X}$. What is particularly interesting here is that the object Q has not to be necessarily contained in $\mathcal{X}$; it will be shown how this specific issue represents one of the main properties of the proposed procedure to extract a knowledge structure out of a set of data.

By Theorem 1 in Huang (1998), $D(\mathcal{X}, \mathcal{Q})$ is minimized if and only if, for every element q_j of $\mathcal{Q}$

$$f_r(A_j = q_j | \mathcal{X}) \geq f_r(A_j = c_{kj} | \mathcal{X}) \text{ for all } j = 1, \dots, m,$$

where: $n_{c_{kj}}$ is the number of objects having the k th category c_{kj} in attribute A_j and $f_r(A_j = c_{kj} | \mathcal{X}) = n_{c_{kj}}/n$ is the relative frequency of category c_{kj} in $\mathcal{X}$. The theorem also states that the mode of a data set is not unique. This issue will play an important role in the following developments of our procedure.

It is now possible to summarize the main issues involved in any of the different algorithms used to implement the k -modes approach (e.g., Ng et al., 2007; San and Huynh, 2004; Chaturvedi et al., 2001; Huang & Ng, 1999; Huang, 1998). First, the algorithm assigns each object to the cluster whose mode is the nearest according to a discrepancy measure and, then, it recomputes the mode on the basis of the objects included in each cluster; second, it reallocates objects into clusters until a certain criterion (e.g., no objects left to reallocate into different clusters, or the discrepancy within each cluster is below a specific value) is satisfied. All the main procedures that apply k -modes follow these fundamental steps and differ one another with respect to specific features, aimed at improving the efficiency of the algorithm or its accuracy.

In the next sections, an adaptation of the standard k -modes algorithm, called k -states, is described. Its accuracy and efficiency are tested through both a simulation study and a practical application to real data. Finally, the perspectives and improvements of the procedure are discussed.

A k -modes approach to knowledge structure extraction

In this section an adaptation of the k -modes algorithm is described, which extracts a knowledge structure from a data set of the responses of N individuals to the dichotomous items in $\mathcal{Q}$. The observed data set is represented by a pair $(\mathcal{R}, F)$, where $\mathcal{R} = 2^{\mathcal{Q}}$ is the power set on the set $\mathcal{Q}$ and $F: \mathcal{R} \rightarrow \mathfrak{R}$ is a function assigning observed frequencies to response patterns. In particular F is such that $F(R) \geq 0$ for all $R \in \mathcal{R}$, and $\sum_{R \in \mathcal{R}} F(R) = N$. Given an initial

knowledge structure $\mathcal{K}_1$ on the set $\mathcal{Q}$, the k -modes algorithm operates in a number $m > 0$ of iterations, each of which consists of the accomplishment of the following two tasks (let $i = 1, 2, \dots, m$ be any iteration number):

- (KM1) given knowledge structure $\mathcal{K}_i$, classify the N observed response patterns into $|\mathcal{K}_i|$ different clusters, each of which is uniquely represented by a knowledge state $K \in \mathcal{K}_i$;
- (KM2) adjust each knowledge state $K \in \mathcal{K}_i$ so that the mean discrepancy between K and the patterns in the class represented by K is minimized. Let $\mathcal{K}_{i+1}$ be the collection of the adjusted knowledge states.

Given an arbitrary knowledge structure $\mathcal{K}$ on $\mathcal{Q}$, the partition of the N observed patterns into the $|\mathcal{K}|$ classes is represented by a function (henceforth called the *partition function*) $f: \mathcal{R} \times \mathcal{K} \rightarrow \mathfrak{R}$, satisfying:

- (C1) $f(R, K) \geq 0$ for all $R \in \mathcal{R}$ and $K \in \mathcal{K}$,
- (C2) $\sum_{K \in \mathcal{K}} f(R, K) = F(R)$ for all $R \in \mathcal{R}$.

The partition function f is interpreted in the following way: given $K \in \mathcal{K}$ and $R \in \mathcal{R}$, f assigns $f(R, K)$ out of $F(R)$ occurrences of response pattern R to the class represented by the knowledge state K . Condition (C2) assures that every occurrence of R is assigned to some state K . It should be observed that the two conditions (C1) and (C2) do not prevent that some of the classes may be empty. That is the possibility that $\sum_{R \in \mathcal{R}} f(R, K) = 0$ for some $K \in \mathcal{K}$ can not be excluded. In this situation, indeed, all the observed response patterns are assigned to a strict subset of $\mathcal{K}$.

Among all possible partition functions for $\mathcal{R}$ and $\mathcal{K}$, the goal is to find one that minimizes some measure of the within-class dissimilarity. A simple dissimilarity measure between $R \in \mathcal{R}$ and $K \in \mathcal{K}$ is the cardinality of their symmetric distance, defined as

$$d(R, K) = |(K \setminus R) \cup (R \setminus K)|.$$

A measure of dissimilarity within the class represented by a knowledge state $K \in \mathcal{K}$ is then obtained as a weighted sum of symmetric distances:

$$D_f(\mathcal{R}, K) = \sum_{R \in \mathcal{R}} f(R, K) d(R, K).$$

The goal is to find a partition function f that minimizes the overall discrepancy

$$D_f(\mathcal{R}, \mathcal{K}) = \sum_{K \in \mathcal{K}} \sum_{R \in \mathcal{R}} f(R, K) d(R, K), \quad (1)$$

which is the sum of the within-class dissimilarities. With the aim of finding such a partition function, in the first place

a lower bound for $D_f(\mathcal{R}, \mathcal{K})$ is established. For this, the following notation will be used. For $R \in \mathcal{R}$, let

$$d_{\min}(R, \mathcal{K}) = \min_{K \in \mathcal{K}} d(R, K)$$

be the minimum distance of response pattern R from the knowledge structure $\mathcal{K}$. Furthermore, define the two collections $\mathcal{K}_R = \{K \in \mathcal{K} : d(R, K) = d_{\min}(R, \mathcal{K})\}$, and $\bar{\mathcal{K}}_R = \mathcal{K} \setminus \mathcal{K}_R$. The collection $\mathcal{K}_R$ contains all states that are at a minimum distance from R , and $\bar{\mathcal{K}}_R$ is its complement in $\mathcal{K}$.

Proposition 1 For any partition function $f : \mathcal{R} \times \mathcal{K} \rightarrow \mathbb{R}$, the following inequality holds true:

$$D_f(\mathcal{R}, \mathcal{K}) \geq \sum_{R \in \mathcal{R}} F(R) d_{\min}(R, \mathcal{K}). \quad (2)$$

Proof By definition, $d(R, K) - d_{\min}(R, \mathcal{K}) \geq 0$ for all $R \in \mathcal{R}$ and $K \in \mathcal{K}$. Hence, since both $f(R, K)$ and $F(R)$ are nonnegative and $\bar{\mathcal{K}}_R \subseteq \mathcal{K}$, it holds that

$$\sum_{K \in \bar{\mathcal{K}}_R} \frac{f(R, K)}{F(R)} (d(R, K) - d_{\min}(R, \mathcal{K})) \geq 0.$$

Expanding:

$$\sum_{K \in \bar{\mathcal{K}}_R} \frac{f(R, K)}{F(R)} d(R, K) - d_{\min}(R, \mathcal{K}) \sum_{K \in \bar{\mathcal{K}}_R} \frac{f(R, K)}{F(R)} \geq 0.$$

But for condition (C2) of the definition of the partition function f , $\sum_{K \in \mathcal{K}} f(R, K)/F(R) = 1$. Moreover $\bar{\mathcal{K}}_R \cup \mathcal{K}_R = \mathcal{K}$, hence

$$\sum_{K \in \bar{\mathcal{K}}_R} \frac{f(R, K)}{F(R)} d(R, K) - d_{\min}(R, \mathcal{K}) \left(1 - \sum_{K \in \mathcal{K}_R} \frac{f(R, K)}{F(R)} \right) \geq 0,$$

from which we obtain

$$d_{\min}(R, \mathcal{K}) \sum_{K \in \mathcal{K}_R} \frac{f(R, K)}{F(R)} + \sum_{K \in \bar{\mathcal{K}}_R} \frac{f(R, K)}{F(R)} d(R, K) \geq d_{\min}(R, \mathcal{K}),$$

but $d_{\min}(R, \mathcal{K}) = d(R, K)$ for all $K \in \mathcal{K}_R$, hence

$$\sum_{K \in \mathcal{K}_R} \frac{f(R, K)}{F(R)} d(R, K) + \sum_{K \in \bar{\mathcal{K}}_R} \frac{f(R, K)}{F(R)} d(R, K) \geq d_{\min}(R, \mathcal{K}),$$

and thus

$$\sum_{K \in \mathcal{K}} f(R, K) d(R, K) \geq F(R) d_{\min}(R, \mathcal{K}).$$

Since, indeed, this last inequality holds for all $R \in \mathcal{R}$, the result follows. $\square$

Call the partition function f a *minimum discrepancy partition function* for $\mathcal{K}$ if it satisfies

$$\sum_{K \in \mathcal{K}_R} f(R, K) = F(R)$$

for all $R \in \mathcal{R}$. A minimum discrepancy partition function is such that $f(R, K) > 0$ if and only if the discrepancy between K and R is minimum. That is, $f(R, K') = 0$ for all $K' \in \bar{\mathcal{K}}_R$.

Proposition 2 The equality

$$D_f(\mathcal{R}, \mathcal{K}) = \sum_{R \in \mathcal{R}} F(R) d_{\min}(R, \mathcal{K})$$

holds true if and only if f is a minimum discrepancy partition function.

Proof The overall discrepancy $D_f(\mathcal{R}, \mathcal{K})$ can be rewritten as

$$D_f(\mathcal{R}, \mathcal{K}) = \sum_{R \in \mathcal{R}} \left(\sum_{K \in \mathcal{K}_R} f(R, K) d(R, K) + \sum_{K \in \bar{\mathcal{K}}_R} f(R, K) d(R, K) \right)$$

Since f is a minimum discrepancy partition function, $f(R, K) = 0$ holds true for all $K \in \bar{\mathcal{K}}_R$, thus

$$D_f(\mathcal{R}, \mathcal{K}) = \sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}_R} f(R, K) d(R, K).$$

Moreover $d(R, K) = d_{\min}(R, \mathcal{K})$ for all $K \in \mathcal{K}_R$, hence

$$D_f(\mathcal{R}, \mathcal{K}) = \sum_{R \in \mathcal{R}} d_{\min}(R, \mathcal{K}) \sum_{K \in \mathcal{K}_R} f(R, K).$$

But $\sum_{K \in \mathcal{K}_R} f(R, K) = F(R)$, therefore

$$D_f(\mathcal{R}, \mathcal{K}) = \sum_{R \in \mathcal{R}} F(R) d_{\min}(R, \mathcal{K}),$$

which completes the proof. $\square$

Thus, any minimum discrepancy partition function f for $\mathcal{K}$ minimizes the overall discrepancy $D_f(\mathcal{R}, \mathcal{K})$. Minimum discrepancy partition functions differ one another with respect to the way they distribute the overall observed frequencies $F(R)$ among the states in $\mathcal{K}_R$. A straightforward special case is represented by the uniform partition function defined by

$$f(R, K) = \begin{cases} F(R)/|\mathcal{K}_R| & \text{if } K \in \mathcal{K}_R \\ 0 & \text{if } K \in \bar{\mathcal{K}}_R. \end{cases}$$

This special case is not new. For instance it can be found in Heller and Wickelmaier (2013) who used it in a different, though related context, for estimating the parameters of a probabilistic knowledge structure by a minimum discrepancy approach.

Concerning step (KM2) of the k -modes algorithm, namely knowledge state adjustment, for $i > 0$ let $\mathcal{K}_i$ be the

knowledge structure obtained at iteration i of the algorithm, f_i be any minimum discrepancy partition function for $\mathcal{K}_i$ and consider any nonempty state $K_i \in \mathcal{K}_i$ such that $K_i \neq Q$ (the empty state and Q are never updated, since they are in $\mathcal{K}_i$ by definition). If $\sum_{R \in \mathcal{R}} f_i(R, K_i) = 0$ (i.e., the class of K_i is empty) then $K_{i+1} = K_i$. Otherwise, a new state K_{i+1} is obtained in the following way: for each item $q \in Q$, the ratio is computed:

$$\theta_{K_i, q} = \frac{\sum_{R \in \mathcal{R}_q} f_i(R, K_i)}{\sum_{R' \in \mathcal{R}} f_i(R', K_i)},$$

where $\mathcal{R}_q = \{R \in \mathcal{R} : q \in R\}$ is the set of all patterns containing q . The ratio $\theta_{K_i, q}$ is the proportion of response patterns containing q , among all those assigned to the class represented by K_i .

Then a decision concerning membership of q to K_{i+1} is made by using the following rule, henceforth called the *state adjustment rule*:

- if $\theta_{K_i, q} > 1/2$ then $q \in K_{i+1}$,
- if $\theta_{K_i, q} < 1/2$ then $q \notin K_{i+1}$,
- if $\theta_{K_i, q} = 1/2$ then $q \in K_{i+1}$ with probability $1/2$.

Indicate with $\mathcal{K}_{i+1}$ the collection of all adjusted knowledge states K_{i+1} and let f_{i+1} be any minimum discrepancy partition function for $\mathcal{K}_{i+1}$. Since the empty set and the full set Q are never adjusted, they are contained in $\mathcal{K}_{i+1}$ and hence $\mathcal{K}_{i+1}$ is a knowledge structure.

Proposition 3 *Let $\mathcal{K}_i$ be any knowledge structure on the set Q , and $f_i : \mathcal{R} \times \mathcal{K}_i \rightarrow \mathbb{R}$ be any minimum discrepancy partition function for $\mathcal{K}_i$. If $\mathcal{K}_{i+1}$ is the knowledge structure obtained from $\mathcal{K}_i$ by an application of the state adjustment rule, then $D_{f_i}(\mathcal{R}, \mathcal{K}_{i+1})$ is minimal in the sense that there is no knowledge structure $\mathcal{K}'$ on Q , with $|\mathcal{K}'| = |\mathcal{K}_{i+1}|$ such that $D_{f_i}(\mathcal{R}, \mathcal{K}') < D_{f_i}(\mathcal{R}, \mathcal{K}_{i+1})$.*

Proof Given any knowledge structure $\mathcal{K}$ on Q , the overall discrepancy $D_{f_i}(\mathcal{R}, \mathcal{K})$ is minimum if the within-class dissimilarity $D_{f_i}(\mathcal{R}, K)$ is minimum for each $K \in \mathcal{K}$. For $R \in \mathcal{R}$, $K \in \mathcal{K}$, $q \in Q$, define the function

$$d_q(R, K) = \begin{cases} 1 & \text{if } q \in K \Delta R, \\ 0 & \text{if } q \in Q \setminus (K \Delta R), \end{cases}$$

with $K \Delta R = (K \setminus R) \cup (R \setminus K)$. Then the within-class dissimilarity can be written as

$$\begin{aligned} D_{f_i}(\mathcal{R}, K) &= \sum_{R \in \mathcal{R}} f_i(R, K) d(R, K) \\ &= \sum_{q \in Q} \sum_{R \in \mathcal{R}} f_i(R, K) d_q(R, K), \end{aligned}$$

and the following decomposition of $D_{f_i}(\mathcal{R}, K)$ is possible:

$$\begin{aligned} D_{f_i}(\mathcal{R}, K) &= \sum_{q \in K} \sum_{R \in \mathcal{R}_q} f_i(R, K) d_q(R, K) \\ &+ \sum_{q \in K} \sum_{R \in \bar{\mathcal{R}}_q} f_i(R, K) d_q(R, K) \\ &+ \sum_{q \in Q \setminus K} \sum_{R \in \mathcal{R}_q} f_i(R, K) d_q(R, K) \\ &+ \sum_{q \in Q \setminus K} \sum_{R \in \bar{\mathcal{R}}_q} f_i(R, K) d_q(R, K), \end{aligned}$$

where $\mathcal{R}_q = \{R \in \mathcal{R} : q \in R\}$, and $\bar{\mathcal{R}}_q = \mathcal{R} \setminus \mathcal{R}_q$. But, for the definition of $d_q(R, K)$, this last decomposition simplifies to

$$D_{f_i}(\mathcal{R}, K) = \sum_{q \in K} \sum_{R \in \bar{\mathcal{R}}_q} f_i(R, K) + \sum_{q \in Q \setminus K} \sum_{R \in \mathcal{R}_q} f_i(R, K).$$

If we define

$$D_{f_i}(\mathcal{R}, K, q) = \begin{cases} \sum_{R \in \bar{\mathcal{R}}_q} f_i(R, K) & \text{if } q \in K, \\ \sum_{R \in \mathcal{R}_q} f_i(R, K) & \text{if } q \in Q \setminus K \end{cases}$$

then, obviously, $D_{f_i}(\mathcal{R}, K) = \sum_{q \in Q} D_{f_i}(\mathcal{R}, K, q)$. Thus, $D_{f_i}(\mathcal{R}, K)$ is minimum if the following condition holds true for every $K' \in 2^Q$ and every $q \in Q$:

$$D_{f_i}(\mathcal{R}, K, q) \leq D_{f_i}(\mathcal{R}, K', q), \quad (3)$$

with $f_i(R, K) = f_i(R, K')$ for all $R \in \mathcal{R}$. If $q \in Q \setminus (K \Delta K')$ then equality obviously holds. If $q \in K \setminus K'$ then Eq. 3 holds true iff

$$\sum_{R \in \bar{\mathcal{R}}_q} f_i(R, K) \leq \sum_{R \in \bar{\mathcal{R}}_q} f_i(R, K').$$

Since $f_i(R, K) = f_i(R, K')$, this last inequality is equivalent to $\theta_{K, q} \geq 1/2$. On the other hand, if $q \in K' \setminus K$ then the opposite inequality is obtained, that is $\theta_{K, q} \leq 1/2$. If K is obtained by the state adjustment rule, then these two conditions are both true by definition. $\square$

Thus in both steps (KM1) and (KM2) of the k -modes algorithm the overall discrepancy $D_f(\mathcal{R}, \mathcal{K})$ is minimized. In particular, at each iteration $i > 0$, in step (KM1) the knowledge structure $\mathcal{K}_i$ is set fixed and the partition function f_{i-1} is replaced by a minimum discrepancy partition function f_i for $\mathcal{K}_i$. This gives $D_{f_i}(\mathcal{R}, \mathcal{K}_i) \leq D_{f_{i-1}}(\mathcal{R}, \mathcal{K}_i)$. In step (KM2) the partition function f_i is set fixed and the knowledge structure $\mathcal{K}_i$ is adjusted so that $D_{f_i}(\mathcal{R}, \mathcal{K}_{i+1}) \leq D_{f_i}(\mathcal{R}, \mathcal{K}_i)$. Thus, the difference $D_{f_{i-1}}(\mathcal{R}, \mathcal{K}_i) - D_{f_i}(\mathcal{R}, \mathcal{K}_{i+1})$ is nonnegative at each iteration $i > 0$. The algorithm terminates when this difference is zero, or below some tolerance value.

It has been stated at the beginning of the section that a partition function cannot prevent empty classes. This

remains true also for a minimum discrepancy partition function. If $d(R, K) > d_{\min}(R, \mathcal{K})$ happens to be true for all $R \in \mathcal{R}$, and a given $K \in \mathcal{K}$, then $\sum_{R \in \mathcal{R}} f(R, K)$ will be zero, meaning that the class of K is empty. Suppose that, for $m > 0$, $\mathcal{K}_m$ is the knowledge structure obtained at the last iteration of the k -modes algorithm. States in $\mathcal{K}_m$ representing empty classes play no role in the classification of the observed response patterns and, for this reason, they can be removed from $\mathcal{K}_m$ with no harm. An obvious exception is represented by the empty set and Q since, by definition, a knowledge structure always contains these two subsets.

Incremental extensions of k -modes

It is well known that k -modes type algorithms are *local minimizers* (see, e.g., Chaturvedi et al., 2001). In the KST context this means that the proposed k -modes algorithm will generally converge to a local minimum of the discrepancy $D_f(\mathcal{R}, \mathcal{K})$ that strictly depends on the input knowledge structure $\mathcal{K}$. Therefore, different input knowledge structures of the same size may lead to different local minima and thus to different solutions.

The question of choosing an input knowledge structure for which the k -modes algorithm will converge to a global minimum is not trivial. Indeed it is not even clear how many knowledge states this input knowledge structure should contain. Obviously, its size should be some number between 2 (the size of the smallest knowledge structure $\mathcal{K}^\perp = \{\emptyset, Q\}$) and $2^{|Q|}$ (the size of the largest one). However, already for the knowledge structure

$$\mathcal{K}^\top = \mathcal{K}^\perp \cup \{R \in \mathcal{R} : F(R) > 0\}$$

one always has $D_f(\mathcal{R}, \mathcal{K}^\top) = 0$, which is a trivial global minimum.

Proposition 4 For any knowledge structure $\mathcal{K} \subseteq 2^Q$, the equality $D_f(\mathcal{R}, \mathcal{K}) = 0$ holds true if and only if $\mathcal{K}^\top \subseteq \mathcal{K}$.

Proof If $\mathcal{K}^\top \subseteq \mathcal{K}$ then $d_{\min}(R, \mathcal{K}) = 0$ for all $R \in \mathcal{R}$ such that $F(R) > 0$. Thus $D_f(\mathcal{R}, \mathcal{K}) = \sum_{R \in \mathcal{R}} F(R) d_{\min}(R, \mathcal{K}) = 0$. If $\mathcal{K}^\top \not\subseteq \mathcal{K}$ then there is $R \in \mathcal{R}$ such that $F(R) > 0$ and $d_{\min}(R, \mathcal{K}) > 0$. Thus $D_f(\mathcal{R}, \mathcal{K})$ cannot be zero. $\square$

In this section an incremental extension of the k -modes algorithm is considered that generates a sequence of $m > 0$ locally optimal knowledge structures $\mathcal{K}_0^*, \mathcal{K}_1^*, \dots, \mathcal{K}_{m-1}^*$ of increasing size, where the smallest structure is $\mathcal{K}_0^* = \{\emptyset, Q\}$.

We shall start by considering the following *trivial incremental extension* of k -modes:

1. let $\mathcal{K}_0 = \{\emptyset, Q\}$ be the *initial knowledge structure*;

2. At each new iteration $j \geq 0$, apply k -modes to $\mathcal{K}_j$, thus obtaining $\mathcal{K}_j^*$;
3. if $D_f(\mathcal{R}, \mathcal{K}_j^*) = 0$ then terminate;
4. else choose a new arbitrary subset $K \in \mathcal{R} \setminus \mathcal{K}_j^*$, form the new knowledge structure $\mathcal{K}_{j+1} = \mathcal{K}_j^* \cup \{K\}$ and repeat from step 2.

Proposition 5 In the trivial incremental extension of k -modes the inequality

$$D_f(\mathcal{R}, \mathcal{K}_{j+1}^*) \leq D_f(\mathcal{R}, \mathcal{K}_j^*)$$

holds at each step $j \geq 0$. In particular, if $K \in \mathcal{R} \setminus \mathcal{K}_j^*$ and $\mathcal{K}_{j+1} = \mathcal{K}_j^* \cup \{K\}$, then strict inequality holds true if there is $R \in \mathcal{R}$ such that $F(R) > 0$ and $d(R, K) < d_{\min}(R, \mathcal{K}_j^*)$.

Proof For $\mathcal{K}_{j+1} = \mathcal{K}_j^* \cup \{K\}$, define the two collections

$$\mathcal{R}_K^+ = \{R \in \mathcal{R} : d(R, K) \geq d_{\min}(R, \mathcal{K}_j^*)\}$$

and

$$\mathcal{R}_K^- = \{R \in \mathcal{R} : d(R, K) < d_{\min}(R, \mathcal{K}_j^*)\}.$$

Since $\mathcal{R}_K^+ \cup \mathcal{R}_K^- = \mathcal{R}$, the discrepancy $D_f(\mathcal{R}, \mathcal{K}_{j+1})$ decomposes as

$$\begin{aligned} D_f(\mathcal{R}, \mathcal{K}_{j+1}) &= \sum_{R \in \mathcal{R}_K^+} F(R) d_{\min}(R, \mathcal{K}_{j+1}) \\ &\quad + \sum_{R \in \mathcal{R}_K^-} F(R) d_{\min}(R, \mathcal{K}_{j+1}) \\ &= \sum_{R \in \mathcal{R}_K^+} F(R) d_{\min}(R, \mathcal{K}_j^* \cup \{K\}) \\ &\quad + \sum_{R \in \mathcal{R}_K^-} F(R) d_{\min}(R, \mathcal{K}_j^* \cup \{K\}). \end{aligned}$$

For the definition of $\mathcal{R}_K^+$, $d_{\min}(R, \mathcal{K}_j^* \cup \{K\}) = d_{\min}(R, \mathcal{K}_j^*)$ for every $R \in \mathcal{R}_K^+$, and for the definition of $\mathcal{R}_K^-$, $d_{\min}(R, \mathcal{K}_j^* \cup \{K\}) = d(R, K)$ for every $R \in \mathcal{R}_K^-$, hence

$$\begin{aligned} D_f(\mathcal{R}, \mathcal{K}_{j+1}) &= \sum_{R \in \mathcal{R}_K^+} F(R) d_{\min}(R, \mathcal{K}_j^*) \\ &\quad + \sum_{R \in \mathcal{R}_K^-} F(R) d(R, K). \end{aligned}$$

Thus we have:

$$\begin{aligned} D_f(\mathcal{R}, \mathcal{K}_j^*) - D_f(\mathcal{R}, \mathcal{K}_{j+1}) &= \sum_{R \in \mathcal{R}_K^-} F(R) d_{\min}(R, \mathcal{K}_j^*) \\ &\quad - \sum_{R \in \mathcal{R}_K^-} F(R) d(R, K) \\ &= \sum_{R \in \mathcal{R}_K^-} F(R) (d_{\min}(R, \mathcal{K}_j^*) \\ &\quad - d(R, K)) \geq 0. \end{aligned}$$

In particular, strict inequality holds whenever there is $R \in \mathcal{R}$ with $F(R) > 0$ and $d(R, K) < d_{\min}(R, \mathcal{K}_j^*)$. Moreover, by Proposition 3, $D_f(\mathcal{R}, \mathcal{K}_{j+1}) \geq D_f(\mathcal{R}, \mathcal{K}_{j+1}^*)$ and therefore $D_f(\mathcal{R}, \mathcal{K}_j^*) > D_f(\mathcal{R}, \mathcal{K}_{j+1}^*)$. $\square$

By Proposition 5, at each step of the trivial incremental extension of k -modes, if $\mathcal{K}^\top \not\subseteq \mathcal{K}_j^*$, then there always exists some new element $K \in \mathcal{R} \setminus \mathcal{K}_j^*$ that improves the overall discrepancy D_f in the sense of reducing it by some amount. This implies that there will also exist a “best” element $\hat{K} \in \mathcal{R} \setminus \mathcal{K}_j^*$ such that, for $\mathcal{K}_{j+1} = \mathcal{K}_j^* \cup \{\hat{K}\}$, the difference $D_f(\mathcal{R}, \mathcal{K}_j^*) - D_f(\mathcal{R}, \mathcal{K}_{j+1}^*)$ is the largest possible. We call $\hat{K} \in \mathcal{R} \setminus \mathcal{K}_j^*$ an *optimal improvement at step j* if it satisfies the inequality

$$D_f(\mathcal{R}, (\mathcal{K}_j^* \cup \{\hat{K}\})^*) \leq D_f(\mathcal{R}, (\mathcal{K}_j^* \cup \{R\})^*)$$

for all $R \in \mathcal{R} \setminus \mathcal{K}_j^*$, where $(\mathcal{K}_j^* \cup \{R\})^*$ is the structure obtained by an application of k -modes to $\mathcal{K}_j^* \cup \{R\}$. An improved version of the incremental extension of k -modes is thus:

1. let $\mathcal{K}_0 = \{\emptyset, Q\}$ be the initial knowledge structure;
2. At each new iteration $j \geq 0$, apply k -modes to $\mathcal{K}_j$, thus obtaining $\mathcal{K}_j^*$;
3. if $D_f(\mathcal{R}, \mathcal{K}_j^*) = 0$ then terminate;
4. else choose an *optimal improvement* $\hat{K} \in \mathcal{R} \setminus \mathcal{K}_j^*$, form the new knowledge structure $\mathcal{K}_{j+1} = \mathcal{K}_j^* \cup \{\hat{K}\}$ and repeat from step 2.

At every single step j of this algorithm a new state $\hat{K} \in \mathcal{R} \setminus \mathcal{K}_j^*$ is added that maximizes the difference

$$D_f(\mathcal{R}, \mathcal{K}_j^*) - D_f(\mathcal{R}, \mathcal{K}_{j+1}^*).$$

Furthermore, since the overall discrepancy $D_f(\mathcal{R}, \mathcal{K}_j^*)$ strictly decreases as $|\mathcal{K}_j^*|$ increases, there must be a step $m > 0$ such that $\mathcal{K}^\top \subseteq \mathcal{K}_m^*$, and $D_f(\mathcal{R}, \mathcal{K}_m^*) = 0$.

When the set of items is not small (say with more than 15 items) the algorithm described above could become rather expensive from a computational point of view, since at every iteration it requires a search in the collection $\mathcal{R} \setminus \mathcal{K}_j^*$. To improve efficiency of the algorithm, it could be useful to restrict the search to some smaller subset of $\mathcal{R} \setminus \mathcal{K}_j^*$. The following corollary states that, if the search is confined to the collection $\mathcal{K}^\top \setminus \mathcal{K}_j^*$ of observed response patterns, then improvement is guaranteed, although it is not known whether it will be optimal.

Corollary 1 *In the trivial incremental extension of k -modes, if $\mathcal{K}_{j+1} = \mathcal{K}_j^* \cup \{K\}$, with $K \in \mathcal{K}^\top \setminus \mathcal{K}_j^*$ then strict inequality $D_f(\mathcal{R}, \mathcal{K}_{j+1}^*) < D_f(\mathcal{R}, \mathcal{K}_j^*)$ holds true.*

Proof From $K \notin \mathcal{K}_j^*$ it follows that $d_{\min}(K, \mathcal{K}_j^*) > 0$. Since K is such that $K \in \mathcal{R}$, $F(K) > 0$ and $d(K, K) =$

$0 < d_{\min}(K, \mathcal{K}_j^*)$, the result immediately follows from Proposition 5. $\square$

Selecting a “best” knowledge structure

Having available the whole set $\{\mathcal{K}_0^*, \mathcal{K}_1^*, \dots, \mathcal{K}_{m-1}^*\}$ of locally optimal knowledge structures of increasing size, the question is now to select a “best” one. If the observed data $(\mathcal{R}, F)$ had been generated by some true, though unknown, knowledge structure $\mathcal{K}_{\text{true}}$ through some probabilistic process (the basic local independence model – BLIM – described by Falmagne and Doignon (1988a) is an example of one such process), then one could seek to find the structure $\mathcal{K}_j^*$ that best approximates the true knowledge structure $\mathcal{K}_{\text{true}}$. Standard model selection criteria like, for instance, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), or still other criteria exist for this purpose. However, to be applicable, they all require a probabilistic framework, which is not established here. It would certainly be possible to provide one: the BLIM itself is an example. Nonetheless this route will not be pursued here for three different reasons.

First, both the AIC and the BIC tend to perform poorly when the sample size is “small” compared to number of parameters (Claeskens & Hjort, 2008; Giraud, 2014). In many probabilistic models for knowledge structures the number of parameters is proportional to the size of the knowledge structure, which could be very large in concrete applications (e.g., thousands of states) even with a moderate number of items (e.g., 20). In this situation, even 1000 would be “small” as a sample size.

The second reason is the efficiency of the extraction procedure, which would be heavily affected by the need of estimating (and re-estimating many times, if local maxima are an issue) model parameters for each of the competing knowledge structures $\mathcal{K}_j^*$.

Thirdly, the choice of any family of parametric models would put, on top of the assumptions of the proposed procedure, all the assumptions of the chosen family. This would make the procedure dependent on the family of parametric models that one chooses. Since the purpose is selecting a best model, the selection process would unavoidably be shaped by the chosen family.

It has been shown in the previous section that the discrepancy $D_f(\mathcal{R}, \mathcal{K}_j^*)$ strictly decreases as $|\mathcal{K}_j^*|$ increases. Given this, a trade-off exists between discrepancy (fit to data) on the one side, and number of knowledge states (model complexity) on the other. Therefore, adhering to a parsimony principle, we can aim at selecting the knowledge structure that displays the “best trade-off” between size $|\mathcal{K}_j^*|$ and discrepancy $D_f(\mathcal{R}, \mathcal{K}_j^*)$. Clearly, this approach comes with no guarantees that the selected model will be the one that best approximates any true knowledge structure. Albeit, the

simulation study described in “[Simulation study](#)” is aimed at investigating this issue systematically, in a number of different conditions.

Considering again the incremental extension of k -modes, the discrepancy decrease at step $j \geq 0$ is

$$s_j = D_f(\mathcal{R}, \mathcal{K}_j^*) - D_f(\mathcal{R}, \mathcal{K}_{j+1}^*),$$

and the average decrease is

$$\begin{aligned}\bar{s} &= \frac{1}{m} \sum_{j=0}^{m-1} s_j = \frac{1}{m} (D_f(\mathcal{R}, \mathcal{K}_0^*) - D_f(\mathcal{R}, \mathcal{K}_{m-1}^*)) \\ &= \frac{1}{m} D_f(\mathcal{R}, \mathcal{K}_0^*).\end{aligned}$$

For $J = \{0, 1, \dots, m-1\}$, consider a criterion that selects the knowledge structure $\mathcal{K}_b^*$ such that

$$b = \min\{j \in J : s_j < \bar{s}\},$$

that is the first structure in the sequence $\mathcal{K}_0^*, \mathcal{K}_1^*, \dots, \mathcal{K}_{m-1}^*$ for which the discrepancy decrease is less than the average decrease. Thus $\mathcal{K}_b^*$ is the first knowledge structure for which the inequality

$$D_f(\mathcal{R}, \mathcal{K}_b^*) - D_f(\mathcal{R}, \mathcal{K}_{b+1}^*) < \bar{s}$$

holds true. If the differences s_j are decreasing (i.e., $s_{j+1} \leq s_j$ for all j) then b is such that $s_i \geq \bar{s}$ for all $i < b$ and $s_j < \bar{s}$ for all $j \geq b$. That is, b separates all knowledge structures for which the discrepancy decrease is faster than average from those for which the decrease is slower than average.

If the discrepancy $D_f(\mathcal{R}, \mathcal{K}_j^*)$ is represented in the plane as a function of the number j of iterations (and hence, indirectly, as a function of the number $|\mathcal{K}_j^*|$ of states), then a kind of “scree plot” similar to the one displayed in Fig. 1, panel (A), can be obtained. If the condition $s_{j+1} \leq s_j$ is respected for all j then the resulting set of points $p_j = (j, D_f(\mathcal{R}, \mathcal{K}_j^*))$ is convex, and the separation described above is possible.

However, in practical applications of the procedure the condition $s_{j+1} \leq s_j$ could not hold for some iteration j , as shown in Fig. 1, panel (B) and therefore the set of points is no longer convex. Points not respecting convexity are removed from the analysis by the following iterative procedure: Let $P = \{(x_0, y_0), (x_1, y_1), \dots, (x_{m-1}, y_{m-1})\}$ denote the initial set of points. At each new iteration of the procedure let $M = \emptyset$ and, for each point $(x_i, y_i) \in P$, if $0 < i < m-1$ and

$$\frac{y_i - y_{i-1}}{x_i - x_{i-1}} > \frac{y_{i+1} - y_i}{x_{i+1} - x_i}$$

then add (x_i, y_i) to M . Once all points (x_i, y_i) have been evaluated this way, replace P by $P \setminus M$. The whole procedure is then repeated with the updated version of P until M is empty.

Fig. 1 Discrepancy decrease as a function of number of states. In panel (a) the condition $s_{j+1} \leq s_j$ is respected for all j , and therefore the resulting set of points is convex. In panel (b) this condition is not respected at $j = 2$

Simulation study

In this section a simulation study is described in which a comparison among three different procedures for data-driven knowledge structure construction was carried out. In particular, the aim was to compare the performance of k -states with *Inductive Item Three Analysis* (IITA; Schrepp 2003) and the procedure proposed by Schrepp (1999a) (in the sequel we refer to this last procedure as the *app-based* procedure).

The three procedures were compared with respect to their capability of recovering a “true” knowledge structure (*goodness of recovery*) from a given data set. In simulating the data set, different conditions were considered, in which three different variables were manipulated: number of items, number of knowledge states and sample size. All data sets were generated by an application of the basic local independence model (BLIM; Falmagne and Doignon, 1988a, b). The BLIM is a probabilistic model for the empirical validation of knowledge structures whose properties have been thoroughly investigated in a number of different studies (de Chiusole et al., 2013, 2015; Doignon & Falmagne, 1999; Falmagne & Doignon, 2011). It assumes the existence of a probability distribution on the knowledge structure $\mathcal{K}$. Furthermore, in this model, the relationship

between the response patterns R and the knowledge state K of a student is given by the unrestricted latent class model

$$P(R) = \sum_{K \in \mathcal{K}} P(R|K) \pi_K,$$

where $P(R)$ is the probability of sampling a student whose response pattern is R , $P(R|K)$ is the conditional probability of observing response pattern R given that the knowledge state is K and π_K is the probability of K . Under the assumption that item responses are locally independent given the knowledge states, for any response pattern R , and any knowledge state $K \in \mathcal{K}$, the conditional probability $P(R|K)$ takes on the form

$$P(R|K) = \left[\prod_{q \in K \setminus R} \beta_q \right] \left[\prod_{q \in K \cap R} (1 - \beta_q) \right] \times \left[\prod_{q \in R \setminus K} \eta_q \right] \left[\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right],$$

where $\beta_q, \eta_q \in [0, 1)$ are called the *careless error probability* and the *lucky guess probability* of item $q \in Q$, respectively.

Having available a knowledge structure, the response patterns of N students can be simulated by using the BLIM. First, a knowledge states K is sampled from the structure with probability π_K . Then, for every item $q \in Q$, random lucky guesses and careless errors are produced with probabilities β_q and η_q .

Before presenting the simulation design, a brief description of the two procedures proposed by Schrepp are given.

Competing procedures

The two procedures proposed by Schrepp (1999a, 2003) are data-driven methodologies for generating a knowledge structure from a set of observed data. These methods share the goal to derive a knowledge structure from data, but differ in the algorithms and in the assumptions they define to reach this goal.

The former (Schrepp, 1999a) consists of an algorithm based on the assumption that the response patterns $R \in \mathcal{R}$ are generated from a set $\mathcal{K}$ of true states by random error, exactly as in the BLIM, but with the following two differences:

- (i) β_q and η_q probabilities are assumed to be equal across items;
- (ii) the probability distribution on the knowledge states is uniform.

The basic mechanism of the algorithm consists of classifying as “states” all response patterns $R \in \mathcal{R}$ having an observed frequency $F(R)$ greater than some cutoff $L > 0$,

and as “non-states” all the other response patterns. Thus, the algorithm is based on the following three steps:

1. sorting the observed response patterns from the one having the highest observed frequency $F(R)$ to the one having the smallest frequency;
2. computing the $app(\mathcal{K}_L, \mathcal{R})$ distance for every possible value of the cutoff L , where:

$$app(\mathcal{K}_L, \mathcal{R}) = \sum_{R \in \mathcal{R}} \frac{(F(R) - F_L(R))^2}{|\mathcal{R}|};$$

$F_L(R)$ is the expected frequency of response pattern R obtained by an application of the BLIM equations to the knowledge structure $\mathcal{K}_L$ (with the restrictions (i) and (ii) described above); $\mathcal{R} = 2^Q$ is the power set on the set Q of items; $\mathcal{K}_L$ is a knowledge structure obtained by collecting all response patterns having an observed frequency greater than L . The smaller the app the better the approximation of the model to the data;

3. selecting the $\mathcal{K}_L$ for which the app is the smallest one.

In the sequel we will refer to this procedure as the *app*-based procedure.

The latter methodology is IITA (Schrepp, 2003). This procedure was developed out of ITA (Bart & Krus, 1973; Van Leeuwe, 1974), and it is aimed at uncovering the logical implications among the items in Q . Such logical implications form a quasi-order (i.e., a reflexive and transitive binary relation) on the set of items. The first crucial assumption of IITA is that the true structure is a quasi-ordinal knowledge space, that is a knowledge structure closed under both union and intersection.

In real data sets some noise is always present, thus, even if a specific implication between two items q and r holds, a certain number b_{qr} of counterexamples of this implication will potentially be included in the response patterns. Let L be the number of counterexamples of an implication from q to r observed in a sample of size m . The main task of IITA is to define quasi-orders $\sqsubseteq_L$ ($L = 0, 1, \dots, m$) for Q . For instance, the relation $q \sqsubseteq_0 r$ involves all those item pairs for which $b_{qr} = 0$. One of the core issues in IITA is the computation of the expected number of counterexamples b_{qr}^* , which has to take into account the estimate of a random error probability γ_{qL} for each item $q \in Q$ and each number L of counterexamples. A fundamental assumption is that $\gamma_{qL} = \gamma_L$ is constant across items.

In the standard formula used by Schrepp (2003), whenever a implication $q \sqsubseteq_L r$ holds true, the expected number of counterexamples is $b_{qr}^* = \gamma_L p_q m$, where p_q is the relative frequency of a correct response to item q . On the contrary, whenever $q \not\sqsubseteq_L r$, no dependency is assumed between items, thus the number of expected counterexamples is $b_{qr}^* = (1 - p_q) p_r m (1 - \gamma_L)$. After generating all the relations $\sqsubseteq_L$, the one that best fits the data is selected. As a

goodness of fit index the *diff* coefficient, defined as follows, is used:

$$\text{diff}(\subseteq_L, \mathcal{R}) = \sum_{q \neq r} \frac{(b_{qr} - b_{qr}^*)^2}{|Q|(|Q| - 1)}.$$

In the present simulation study we used the modified version of IITA presented by Sargin and Ünlü (2009). In their article, the authors highlight some of the main criticisms of IITA and propose some solutions for correctly addressing the computation of b_{qr}^* in the case in which $q \not\subseteq_L r$. More specifically, they recognize that when $q \not\subseteq_L r$ two different configurations can hold: in the first one $r \not\subseteq_L q$; in the second one $r \subseteq_L q$. In the former case independence holds between q and r , thus $b_{qr}^* = (1 - p_q)p_r m$. In the latter case independence cannot be assumed. Authors present the following correction of the estimate for this case: $b_{qr}^* = (p_r - p_q(1 - \gamma_L))m$ (Sargin & Ünlü, 2009). Furthermore, they introduced an improvement of the procedure that minimizes the *diff* coefficient with respect to the error probability γ_L .

These modifications have been implemented into the DAKS package for R (Ünlü & Sargin, 2010), that has been used in this simulation study.

Simulation Design

Table 1 shows the simulation design of the study.

Ten different conditions were considered in which the following three variables were manipulated:

- number of items $q \in Q$: 10 or 15;
- number of knowledge states $K \in \mathcal{K}$: 35, 70, 150 or 300;
- sample size N : 500, 1000, 2000, 3000 or 4000.

It is difficult to establish an ideal or average number of knowledge states for a given number of items (it could

Table 1 Simulation design of the study. In column 1, the ten different conditions are displayed.

Condition	$ Q $	$ \mathcal{K} $	N
1	10	35	500
2	10	35	1000
3	10	70	500
4	10	70	1000
5	15	150	1000
6	15	150	2000
7	15	300	1000
8	15	300	2000
9	15	300	3000
10	15	300	4000

The number of items $|Q|$, the number of knowledge states $|\mathcal{K}|$ and the sample sizes N are listed in columns 2, 3, and 4, respectively

vary from application to application). It is only reasonable to expect that this number increases with the number of items. In our simulation design we used: 35 or 70 states with 10 items and 150 or 300 states with 15 items. Concerning the sample size, it seems reasonable to expect that a large number of observations is needed for extracting a knowledge structure containing a large number of states. The following sizes were used: $N \in \{500, 1000\}$ with less than 150 states, $N \in \{1000, 2000\}$ with 150 states, and $N \in \{1000, 2000, 3000, 4000\}$ with 300 states.

In the whole, 4 different knowledge structures were considered: (1) $\mathcal{K}_1$ for conditions 1 and 2; (2) $\mathcal{K}_2$ for conditions 3 and 4; (3) $\mathcal{K}_3$ for conditions 5 and 6; (4) $\mathcal{K}_4$ for conditions 7, 8, 9 and 10. These four knowledge structures were obtained by computing $\{\emptyset, Q\} \cup \mathcal{P}$, where $\mathcal{P}$ was generated at random, using a sampling without replacement on the collection $2^Q \setminus \{\emptyset, Q\}$.

In each condition, 100 simulated data sets of size N were generated by the BLIM, in which the β_q and η_q parameters of the items were generated by using a uniform distribution in the interval $(0, .1]$. The probabilities π_K of the knowledge states $K \in \mathcal{K}_n$, where $n \in \{1, 2, 3, 4\}$, were generated by using a uniform distribution in the interval $[.4, .6]$, and then they were normalized to sum up to 1. The β_q and η_q parameter values and the knowledge state probabilities were kept constant across simulation conditions using the same knowledge structure. In the whole $10 \times 100 = 1,000$ data sets were generated.

Comparison among the three procedures: performance indexes

In order to compare the true knowledge structure with the knowledge structure extracted from the data by the particular procedure, five different performance indexes were considered. All of them evaluate the goodness of recovery of the procedures.

1. The *true positive rate* (TPR) is the proportion of true knowledge states $K \in \mathcal{K}$ belonging to the extracted knowledge structure $\mathcal{K}_e$. Formally:

$$\text{TPR} = \frac{|\mathcal{K}_e \cap \mathcal{K}|}{|\mathcal{K}|} \quad (4)$$

2. The *false positive rate* (FPR) is the proportion of knowledge states $K \in \mathcal{K}_e$ not belonging to the true knowledge structure $\mathcal{K}$. Formally:

$$\text{FPR} = \frac{|\mathcal{K}_e \setminus \mathcal{K}|}{|\mathcal{K}_e|} \quad (5)$$

3. The average discrepancy between $\mathcal{K}_e$ and $\mathcal{K}$:

$$D(\mathcal{K}_e, \mathcal{K}) = \frac{1}{|\mathcal{K}_e|} \sum_{K \in \mathcal{K}_e} d_{\min}(K, \mathcal{K}), \quad (6)$$

Table 2 Comparison among the performance of k -states (*top panel*), *app*-based (*mid panel*) and IITA (*bottom panel*) procedures in their goodness of recovery.

Condition	$ R $	$ \mathcal{K}_e $	TPR	FPR	$D(\mathcal{K}_e, \mathcal{K})$	$D(\mathcal{K}, \mathcal{K}_e)$	κ
<i>k</i> -states							
1	197.71	33.34	0.93	0.02	0.08	0.02	0.96
2	293.41	35.94	1.00	0.03	< 0.01	0.03	0.99
3	237.95	44.79	0.57	0.11	0.52	0.11	0.68
4	354.45	58.05	0.79	0.05	0.23	0.05	0.86
5	595.09	110.79	0.68	0.08	0.61	0.09	0.78
6	985.80	142.28	0.94	0.01	0.09	0.01	0.96
7	712.66	127.40	0.32	0.25	1.40	0.27	0.44
8	1181.50	196.66	0.58	0.12	0.73	0.12	0.69
9	1577.40	242.64	0.76	0.06	0.37	0.06	0.84
10	1933.46	270.07	0.88	0.03	0.17	0.03	0.92
<i>app</i> -based							
Condition	$ R $	$ \mathcal{K}_e $	TPR	FPR	$D(\mathcal{K}_e, \mathcal{K})$	$D(\mathcal{K}, \mathcal{K}_e)$	κ
1	197.71	70.81	0.99	0.50	0.01	0.52	0.65
2	293.41	62.23	1.00	0.41	< 0.01	0.41	0.72
3	237.95	101.17	0.91	0.34	0.11	0.36	0.74
4	354.45	141.42	0.99	0.49	0.01	0.51	0.63
5	595.09	561.02	0.95	0.75	0.06	0.99	0.40
6	985.80	519.94	1.00	0.71	< 0.01	0.75	0.44
7	712.66	676.19	0.78	0.65	0.29	0.84	0.47
8	1181.50	1181.50	0.96	0.76	0.04	0.98	0.38
9	1577.40	1577.40	0.99	0.81	0.01	1.06	0.31
10	1933.46	609.29	0.99	0.51	0.02	0.55	0.65
IITA							
Condition	$ R $	$ \mathcal{K}_e $	TPR	FPR	$D(\mathcal{K}_e, \mathcal{K})$	$D(\mathcal{K}, \mathcal{K}_e)$	κ
1	197.71	186.22	0.39	0.93	0.82	1.43	0.07
2	293.41	190.14	0.41	0.92	0.79	1.42	0.07
3	237.95	144.63	0.20	0.90	1.20	1.25	0.04
4	354.45	152.68	0.21	0.90	1.14	1.25	0.04
5	595.09	4278.23	0.23	0.99	1.44	2.32	0.01
6	985.80	4226.91	0.22	0.99	1.47	2.31	0.01
7	712.66	2456.79	0.11	0.99	1.88	2.01	0.01
8	1181.50	2371.85	0.11	0.99	1.81	2.00	0.01
9	1577.40	2347.10	0.11	0.99	1.77	1.99	0.01
10	1933.46	2322.70	0.11	0.99	1.73	1.99	0.01

Column 1 displays the number of the ten conditions and column two displays the average number of the simulated response patterns; the last five columns display the mean of the performance indexes obtained across the 100 simulated samples. Going from left to right, the following indexes are reported: the cardinality of the extracted knowledge structure $\mathcal{K}_e$, the true positive rate (TPR), the false positive rate (FPR), the $D(\mathcal{K}_e, \mathcal{K})$ and $D(\mathcal{K}, \mathcal{K}_e)$ discrepancies and Cohen's κ

where $d_{\min}(K, \mathcal{K}) = \min_{K' \in \mathcal{K}} d(K, K')$ is the minimum discrepancy between the knowledge state $K \in \mathcal{K}_e$ and the true knowledge structure $\mathcal{K}$.

4. The average discrepancy between K and $\mathcal{K}_e$:

$$D(\mathcal{K}, \mathcal{K}_e) = \frac{1}{|\mathcal{K}|} \sum_{K \in \mathcal{K}} d - \min(K, \mathcal{K}_e), \quad (7)$$

where $d_{\min}(K, \mathcal{K}_e) = \min_{K' \in \mathcal{K}_e} d(K, K')$ is the minimum symmetric distance between the knowledge state $K \in \mathcal{K}$ and the knowledge structure $\mathcal{K}_e$.

5. The *Cohen's* κ , computed for the following observed frequencies:

- number of positive agreements: $|\mathcal{K}_e \cap \mathcal{K}|$;
- number of negative agreements: $2^{|\mathcal{Q}|} - |\mathcal{K} \cup \mathcal{K}_e|$;

- number of false positives: $|\mathcal{K}_e \setminus \mathcal{K}|$;
- number of false negatives: $|\mathcal{K} \setminus \mathcal{K}_e|$.

The number of false positives plus the number of false negatives gives the total number of observed disagreements.

The five indexes were computed for each of the 100 simulated samples in each of the ten conditions displayed in Table 1. Then, an average value across of the 100 samples was obtained for each index in each condition.

Results

Table 2 displays the results concerning the goodness of recovery of the k -states (top panel) and the other two competing (mid and bottom panel) procedures. The first two columns of the table display the number of the simulation conditions and the average number of simulated response patterns respectively, whereas columns 3 to 8 display the mean of the performance indexes computed for the knowledge structure extracted by the three procedures.

It can be observed that the cardinalities of the extracted knowledge structures for the *app*-based and the IITA procedures is systematically greater than both $\mathcal{K}_e$ extracted by k -states and the true knowledge structure. This happens irrespectively of the number of items, the sample and knowledge structure sizes. Except for conditions in which the number of items was 10, the cardinality of the $\mathcal{K}_e$ extracted by the *app*-based procedure approaches the number of the simulated response patterns. This means that almost all the response patterns are included in $\mathcal{K}_e$. Whereas, knowledge structures extracted by IITA far outweigh the cardinality of the true structures, doubling in most cases the number of the simulated response patterns.

On the one hand, this result has a positive effect on the TPR index for the *app*-based and k -states procedures: It is always higher for the *app*-based procedure, meaning that it extracts a greater number of true knowledge states. On the other hand, this result has a negative effect on the FPR index: It is higher for *app*-based procedure, meaning that its extracted knowledge structures contain many false states (in some conditions more than 70 %). Conversely, k -states seems to be more parsimonious, preferring the choice of few but true knowledge states. In fact, the FPR index of k -states is systematically the lowest one, with percentages that are much smaller than those of the other two procedures. Concerning IITA, in all conditions the TPR index is much smaller than the FPR, that in many cases approaches 1.

All these results are reflected by the two discrepancies $D(\mathcal{K}_e, \mathcal{K})$ and $D(\mathcal{K}, \mathcal{K}_e)$: The former is smaller for the *app*-based procedure, the latter is smaller for k -states. Furthermore, irrespectively of the considered extraction method, the higher the TPR, the lower the discrepancy $D(\mathcal{K}_e, \mathcal{K})$, whereas the higher the FPR, the higher $D(\mathcal{K}, \mathcal{K}_e)$.

To provide a more synthetic index for comparing the accuracy of the two methods, we computed also Cohen's κ (last column of Table 2), which takes concurrently into account both the information on TPR and FPR. It can be noted that, in 8 out of 10 conditions, Cohen's κ is higher for the knowledge structures extracted by k -states, approaching in some conditions (1, 2 and 6) the upper bound 1.

The worsening of the performance of k -states in conditions number 3 and 7 could be due to an inadequate sample size compared to the number of states in the structure. With 10 items and 70 knowledge states, a sample size of 500 might be inadequate for recovering the true knowledge structure by using k -states. Indeed, by increasing the sample size to 1000 (condition 4), also k -states' performance improves. The same reasoning applies to condition 7, where a sample size of 1000 seems to be inadequate with 15 items and 150 states.

These results suggest that increasing the amount of empirical evidence positively affects the k -states' performance: the TPR percentage increases while that of the FPR decreases in most cases. On the contrary, in the *app*-based procedure the FPR is systematically very high irrespectively of the sample size. Concluding, an increase of the sample size improves k -states' performance, but paradoxically, it worsens that of the *app*-based procedure. Concerning IITA, the increase of sample size seems having no effect on its performance.

Empirical application

In order to apply k -states to real data, an empirical application was carried out. The aim was to (1) apply all three competing procedures to a real data set, thus obtaining three (possibly) different structures; (2) fit the BLIM to the data for each of the extracted structures.

Methods

The application was carried out starting from a set of answers to the reduced form of the Maudsley Obsessional-Compulsive Questionnaire (MOCQ-R; Sanavio & Vidotto, 1985), a questionnaire investigating the obsessive and compulsive symptoms included in the wide spectrum clinical assessment battery CBA 2.0 (Sanavio et al., 2008). The questionnaire presents a dichotomous answer format and it is composed of 21 items divided into three subscales investigating three of the main dimensions of the Obsessive Compulsive Disorder (OCD): "Checking", "Cleaning" and "Doubting-Ruminating". More specifically, the application was focused on the first two subscales containing a total of 16 items. The sample was composed of 4412 individuals and it was used in previous research (Spoto

et al. 2010, 2012); the questionnaire was administered during a wider assessment procedure. Participants signed the informed consent and were asked to answer to all the items of the questionnaire. No time limit was imposed. A number of 4297 out of 4412 filled questionnaires were used for the analysis, while incomplete questionnaires were excluded. The three procedures were applied to these data in order to extract a structure for the 16 items.

In order to test the goodness of fit of the obtained structures, a sample of 59 patients with diagnosis of OCD (formulated by experts in Cognitive Behavioral Therapy) was used. Patients filled the MOCQ-R during the assessment phase of their treatment. The BLIM was then fitted to the clinical data set for each of the three extracted structures. This analysis was aimed at comparing the goodness of fit of the structures obtained through each procedure.

The goodness-of-fit of the BLIM was tested by the Pearson Chi-square statistic. A parametric bootstrap procedure (Efron, 1979) over 1000 replications was used to compute the p -value of the Chi-square. This is necessary because the approximation to the asymptotic distribution of the Chi-square statistic lacks accuracy for large and sparse data matrices and this was the case of the present empirical application.

Parametric bootstrap was performed in the following way: (i) The parameter estimates obtained by fitting the model to the data were used to generate 1000 data sets of the same size of the observed sample ($N = 59$); (ii) the model was fitted to each of the 1000 simulated data sets; and (iii) the proportion of replications in which the model obtained a Chi-square greater than the observed one was the bootstrapped p -value.

Results

The *app*-based procedure extracted a structure containing only the empty set and the total set: An overly simple model. It is likely that the restrictions at the basis of this procedure are too strong, leading the *app* selection criterion to be excessively conservative. Needless to say, the fit of the BLIM to data led to a strong rejection.

Concerning IITA, it extracted a structure of 103 states, but the p -value of the BLIM's Chi-square computed for this structure was $p = .07$ (which, considering a first type error probability $\alpha = .10$, leads to a rejection).

Finally, k -states extracted a structure composed of 246 states. Identifiability of the BLIM for the extracted structure was tested by the BLIMIT function (Stefanutti et al., 2012). No identifiability issues were found. The p -value of the BLIM's Chi-square, computed for this structure, was $p = .23$, meaning that the model predicts the data quite well. Also the values of the error parameter estimates $\hat{\beta}_q$ and $\hat{\eta}_q$ of the items suggest a very good fit of the model to the

data. Indeed, the following average values across items were obtained: $\bar{\beta}_q = .11$ (SD = .09) and $\bar{\eta}_q = .13$ (SD = 0.16).

Discussion

In the present paper a new data-driven procedure to build knowledge structures, called k -states, was presented. The development of the procedure drawn upon the area of data mining and, in particular, to the k -modes clustering (Chaturvedi et al., 2001; Huang, 1998). The proposed algorithm is an incremental extension of k -modes that generates a sequence of locally optimal knowledge structures of increasing size, among which a “best” model is selected.

In order to test the applicability of the k -states algorithm, a simulation study and an empirical application were carried out. In the former study the aim was to compare the performance of k -states with that of the *app*-based (Schrepp, 1999a) and the IITA (Schrepp, 2003) procedures, in different simulation conditions. In the comparison of the knowledge structures extracted by the three procedures with those used for generating the data, k -states performed better in most cases. Despite k -states performs quite well in all the simulation conditions, it extracts knowledge structures with cardinality systematically smaller than that of the “true” structure. This suggests that the selection criterion is rather conservative and needs improvements. Further studies should investigate other types of selection criteria that allow to improve the proportion of “true” states, while keeping low the proportion of “false” states contained in the resulting structure.

Concerning the empirical application, each of the three procedures extracted a structure from an existing data set of 4297 respondents to 16 items of the MOCQ-R (MOCQ-R; Sanavio & Vidotto, 1985). The only structure for which the BLIM obtained an acceptable fit was the one extracted by k -states.

Concluding, the strengths of k -states, compared to the other procedures, can be summarized as follows: (1) unlike IITA and IITA, k -states does not assume any restriction on the properties of the knowledge structure; (2) Unlike the *app*-based procedure, it does not require that the collection of knowledge states is a subset of the observed response patterns; (3) In selecting the states that have to be included in the structure, k -states seems to be more parsimonious than the *app*-based procedure, preferring few but correct knowledge states.

Acknowledgments The research developed in this article was carried out under the research project CPDA149902, funded by the University of Padua. The authors would like to acknowledge Giorgio Bertolotti and Salvatore Maugeri Foundation for providing the data set used in the empirical example.

References

- Albert, D., & Lukas, J. (Eds.) (1999). *Knowledge Spaces: Theories, Empirical Research, and Applications Knowledge spaces: Theories, empirical research and applications*: Mahwah, NJ: Lawrence Erlbaum Associates.
- Bart, W.M., & Krus, D.J. (1973). An ordering-theoretic method to determine hierarchies among items. *Educational and psychological measurement*.
- Chaturvedi, A., Green, P.E., & Carroll, J.D. (2001). K-modes clustering K-modes clustering. *Journal of Classification*, 18(1), 35–55.
- de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2013). Assessing Parameter Invariance in the BLIM: Bipartition Models Assessing parameter invariance in the BLIM: Bipartition Models. *Psychometrika*, 78(4), 710–724.
- de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). Modeling missing data in knowledge space theory. Modeling missing data in knowledge space theory. *Psychological Methods*, 20(4), 506.
- Claeskens, G., & Hjort, N.L. (2008). *Model selection and model averaging Model selection and model averaging (Vol 330)*. Cambridge University Press, Cambridge.
- Cosyn, E., & Thiéry, N. (2000). A practical procedure to build a knowledge structure A practical procedure to build a knowledge structure. *Journal of mathematical psychology*, 44(3), 383–407.
- Desmarais, M.C., Maluf, A., & Liu, J. (1995). User-expertise modeling with empirically derived probabilistic implication networks User-expertise modeling with empirically derived probabilistic implication networks. *User modeling and user-adapted interaction*, 5(3-4), 283–315.
- Doignon, J.-P. (1994). Knowledge spaces and skill assignments Knowledge spaces and skill assignments. In Fischer, G., & Laming, D. (Eds.) *Contributions to Mathematical Psychology, Psychometrics and Methodology Contributions to mathematical psychology, psychometrics and methodology (p. 111-121)*. New York Springer-Verlag.
- Doignon, J.P., & Falmagne, J.C. (1985). Spaces for the assessment of knowledge Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23, 175–196.
- Doignon, J.P., & Falmagne, J.C. (1999). *Knowledge spaces Knowledge spaces*. New York: Springer.
- Dowling, C.E. (1993). Applying the basis of a knowledge space for controlling the questioning of an expert Applying the basis of a knowledge space for controlling the questioning of an expert. *Journal of Mathematical Psychology*, 37(1), 21–48.
- Efron, B. (1979). Bootstrap methods: Another look at jackknife. *Annual Statistics*, 7(2), 1–2.
- Falmagne, J.C., Albert, D., Doble, C., Eppstein, D., & Hu, X. (2013). *Knowledge spaces: Applications in education Knowledge spaces*. Springer Science & Business Media.
- Falmagne, J.C., & Doignon, J.P. (1988a). A class of stochastic procedures for the assessment of knowledge 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 state of a system A Markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology*, 32, 232–258.
- Falmagne, J.C., & Doignon, J.P. (2011). *Learning spaces Learning spaces*. New York: Springer.
- Giraud, C. (2014). *Introduction to high-dimensional statistics Introduction to high-dimensional statistics (Vol. 138)*: CRC Press.
- Hartigan, J.A., & Wong, M.A. (1979). Algorithm AS 136: A k-means clustering algorithm Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.
- Heller, J., Augustin, T., Hockemeyer, C., Stefanutti, L., & Albert, D. (2013). *Recent Developments in Competence-based Knowledge Space Theory Recent developments in competence-based knowledge space theory. Knowledge Spaces Knowledge spaces (p. 243–286)*: Springer.
- Heller, J., & Wickelmaier, F. (2013). Minimum Discrepancy Estimation in Probabilistic Knowledge Structures Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, 42(4), 49–56.
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, 2(3), 283–304.
- Huang, Z., & Ng, M.K. (1999). A fuzzy k-modes algorithm for clustering categorical data A fuzzy k-modes algorithm for clustering categorical data. *IEEE Transactions on Fuzzy Systems*, 7(4), 446–452.
- Kambouri, M., Koppen, M., Villano, M., & Falmagne, J.C. (1994). Knowledge assessment: Tapping human expertise by the QUERY routine Knowledge assessment: Tapping human expertise by the query routine. *International Journal of Human-Computer Studies*, 40(1), 119–151.
- Koppen, M. (1993). Extracting human expertise for constructing knowledge spaces: An algorithm Extracting human expertise for constructing knowledge spaces: An algorithm. *Journal of mathematical psychology*, 37(1), 1–20.
- Koppen, M., & Doignon, J.P. (1990). How to build a knowledge space by querying an expert How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology*, 34(3), 311–331.
- Lukas, J., & Albert, D. (1993). Knowledge Assessment Based on Skill Assignment and Psychological Task Analysis Knowledge assessment based on skill assignment and psychological task analysis. In Strube, G., & Wender, K. (Eds.) *The Cognitive Psychology of Knowledge The cognitive psychology of knowledge (p. 139-160)*. Amsterdam North-Holland.
- Müller, C.E. (1989). *A procedure for facilitating an experts judgements on a set of rules A procedure for facilitating an experts judgements on a set of rules. Mathematical psychology in progress (pp. 157–170)*: Springer.
- Ng, M.K., Li, M.J., Huang, J.Z., & He, Z. (2007). On the impact of dissimilarity measure in k-modes clustering algorithm On the impact of dissimilarity measure in k-modes clustering algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), 503–507.
- Robusto, E., & Stefanutti, L. (2014). Extracting a knowledge structure from the data by a maximum residuals method Extracting a knowledge structure from the data by a maximum residuals method. TPM: Testing, Psychometrics, Methodology in Applied Psychology.
- San, O. M., & Huynh, N. (2004). An alternative extension of the k-means algorithm for clustering categorical data An alternative extension of the k-means algorithm for clustering categorical data. *International Journal of Applied Mathematics and Computer Science*, 14(2), 241–248.
- Sanavio, E., Bertolotti, G., Michielin, P., Vidotto, G., & Zotti, A. (2008). CBA-2.0 scale primarie: Manuale. Una batteria ad ampio spettro per l'assessment psicologico. Cba-2.0 scale primarie: Manuale. una batteria ad ampio spettro per l'assessment psicologico. Firenze: Organizzazioni Speciali.

- Sanavio, E., & Vidotto, G. (1985). The components of the Maudsley obsessional-compulsive questionnaire The components of the Maudsley obsessional-compulsive questionnaire. *Behaviour Research and Therapy*, 23(6), 659–662.
- Sargin, A., & Ünlü, A. (2009). Inductive item tree analysis: Corrections, improvements, and comparisons Inductive item tree analysis: Corrections, improvements, and comparisons. *Mathematical Social Sciences*, 58(3), 376–392.
- Schrepp, M. (1999a). Extracting Knowledge Structures from Observed Data Extracting Knowledge Structures from Observed Data. *British Journal of Mathematical and Statistical Psychology*, 52, 213–224.
- Schrepp, M. (1999b). On the empirical construction of implications between bi-valued test items On the empirical construction of implications between bi-valued test items. *Mathematical social sciences*, 38(3), 361–375.
- Schrepp, M. (2002). Explorative analysis of empirical data by Boolean analysis of questionnaires. *Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie*.
- Schrepp, M. (2003). A method for the analysis of hierarchical dependencies between items of a questionnaire A method for the analysis of hierarchical dependencies between items of a questionnaire. *Methods of Psychological Research Online*, 19, 43–79.
- Schrepp, M., & Held, T. (1995). A simulation study concerning the effect of errors on the establishment of knowledge spaces by querying experts A simulation study concerning the effect of errors on the establishment of knowledge spaces by querying experts. *Journal of Mathematical Psychology*, 39(4), 376–382.
- Spoto, A., Stefanutti, L., & Vidotto, G. (2010). Knowledge space theory, formal concept analysis, and computerized psychological assessment Knowledge space theory, formal concept analysis, and computerized psychological assessment. *Behavior Research Methods*, 42(1), 342–350.
- Spoto, A., Stefanutti, L., & Vidotto, G. (2012). On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures. *Journal of Mathematical Psychology*, 56(4), 248–255.
- Spoto, A., Stefanutti, L., & Vidotto, G. (2015). An iterative procedure for extracting skill maps from data An iterative procedure for extracting skill maps from data. *Behavior research methods*, 1–13.
- Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E. (2012). Assessing the local identifiability of probabilistic knowledge structures Assessing the local identifiability of probabilistic knowledge structures. *Behaviour Research Methods*, 44, 1197–1211.
- Ünlü, A., & Albert, D. (2004). The Correlational Agreement Coefficient $CA(\leq, D)$ a mathematical analysis of a descriptive goodness-of-fit measure The correlational agreement coefficient $ca(\leq, d)$ a mathematical analysis of a descriptive goodness-of-fit measure. *Mathematical Social Sciences*, 48(3), 281–314.
- Ünlü, A., & Sargin, A. (2010). DAKS: An R package for data analysis methods in knowledge space theory Daks: An R package for data analysis methods in knowledge space theory. *Journal of Statistical Software*, 37(2), 1–31.
- Van Leeuwe, J.F. (1974). Item tree analysis. Item tree analysis. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*.
- Villano, M. (1991). Computerized knowledge assessment: Building the knowledge structure and calibrating the assessment routine. Unpublished doctoral dissertation.