Psychometric data analysis: A size/fit trade-off evaluation procedure for knowledge structures

Ali Ünlü, Waqas Ahmed Malik

Universität Augsburg

Institut für Mathematik

Ali Ünlü, Waqas Ahmed Malik

Psychometric Data Analysis: A Size/Fit Trade-Off Evaluation Procedure for Knowledge Structures

Psychometric data analysis: A size/fit trade-off evaluation procedure for knowledge structures

Ali Ünlü[*], Waqas Ahmed Malik

Department of Computer-Oriented Statistics and Data Analysis, Institute of Mathematics, University of Augsburg, Universitätsstraße 14, D-86159 Augsburg, Germany

ali.uenlue@math.uni-augsburg.de, malikwaqasahmed@gmail.com

February 5, 2008

Abstract   A crucial problem in knowledge space theory, a modern psychological test theory, is the derivation of a realistic knowledge structure representing the organization of knowledge in an information domain and examinee population under reference. Often, one is left with the problem of selecting among candidate competing knowledge structures. This article proposes a measure for the selection among competing knowledge structures. It is derived within an operational framework (prediction paradigm), and is partly based on the unitary method of proportional re-

---
[*] Corresponding author. Phone: +49 821 598 2236; fax: +49 821 598 2200.

duction in predictive error as advocated by the authors Guttman, Goodman, and Kruskal. In particular, this measure is designed to trade off the (descriptive) fit and size of a knowledge structure, which is of high interest in knowledge space theory. The proposed approach is compared with the Correlational Agreement Coefficient, which has been recently discussed for the selection among competing surmise relations. Their performances as selection measures are compared in a simulation study using the fundamental basic local independence model in knowledge space theory.

Key words   Exploratory data analysis; Knowledge space theory; Latent variable model; Psychometrics; Selection measure; Size/fit trade-off

1 Introduction

Knowledge structures and *surmise relations* are mathematical models that belong to the theory of *knowledge spaces* (reviewed in Section 2). *Knowledge space theory* (KST) was introduced by Doignon and Falmagne (1985, 1999), and it has been successfully applied for the computerized, adaptive assessment and training of knowledge; for instance, see the ALEKS

(Assessment and LEarning in Knowledge Spaces) system, a fully automated math tutor on the Internet.[1] KST models have also been applied in such areas as the structuring of hyper-texts, the analysis of organizational workflows, and the modeling of cross-cultural knowledge and inter/intra-cultural value systems.[2]

However, a crucial problem in KST is the derivation of a 'realistic' knowledge structure from empirical data, representing the organization of 'knowledge' in an information domain and examinee population under reference. In this regard, often one has to make a choice among candidate competing knowledge structures. (For instance, Section 5 describes how the candidate competing knowledge structure models may be obtained data-analytically, based on a *modified Item Tree Analysis* procedure. Or, the competing models under consideration may be derived theoretically, based on different psychological theories/postulates.)

---

[1] `http://www.aleks.com`

[2] For a comprehensive list of references on KST and its applications, refer to the web sites at `http://css.uni-graz.at/staff/hockemeyer/kst-bib.html` and `http://css.uni-graz.at/kst.php`, which are maintained by C. Hockemeyer and M. D. Kickmeier-Rust (University of Graz, Austria).

A measure κ is proposed for the evaluation of knowledge structures. It is designed to trade off the (descriptive) fit of a knowledge structure to a given data set and its size. Such a trade-off is of high interest in KST. (For instance, Section 4.5 mentions that this type of trade-off is beneficial for the efficient application of adaptive knowledge assessment procedures. In general, for a knowledge structure of a smaller size only a fewer items have to be answered by an examinee to assess her/his state of knowledge. Of course, the assessment procedure must also be based on a, more or less, 'valid' (data fitting) knowledge structure underlying an examinee's response behavior. In addition, the more states are contained in a knowledge structure (larger size) the smaller are the distances of the observed response patterns to the closest states in the knowledge structure (better fit). Therefore, any such measure must realize a trade-off between the size of a knowledge structure and its fit to the data.) The measure κ is derived within an operational framework (prediction paradigm), partly based on the unitary method of *proportional reduction in predictive error* (Section 4).

The approach to 'model selection' among knowledge structures based on the measure κ is compared with the *Correlational Agreement Coefficient* CA (reviewed in Section 3), which has been recently discussed for the selection among surmise relations. The performances of κ and CA as selection measures are compared in a simulation study using the *basic local independence model*, which is a fundamental finite mixture, latent variable model in KST.

On the structure of this article. Section 2 reviews basic deterministic and probabilistic concepts of KST that are relevant for this work. Section 3 recapitulates the Correlational Agreement Coefficient CA. Section 4 proposes the size/fit trade-off evaluation procedure κ. Section 6 discusses an application of κ and CA to simulated data. This article concludes with a discussion in Section 7, containing a summary and some suggestions for further research.

2 Knowledge space theory

This section starts with a motivating small example which is taken from Falmagne *et al.* (2003), and then briefly reviews some of the basic de-

terministic and probabilistic concepts of KST. For details, the reader is referred to Doignon and Falmagne (1999).

2.1 Example: Elementary Algebra

A natural starting point for a theory of knowledge assessment and training stems from the observation that some pieces of knowledge may imply other pieces of knowledge. In the context of this section, the mastery of some algebra problem may imply the mastery of other problems. Such implications between pieces of knowledge may be used to design efficient computer-based, adaptive knowledge assessment and training procedures (cf. Section 4.5).

Consider the six dichotomous problems in Elementary Algebra:

a. A car travels on the freeway at an average speed of 52 miles per hour. How many miles does it travel in 5 hours and 30 minutes?

b. Using the pencil, mark the point at the coordinates $(1, 3)$.

c. Perform the following multiplication:

$$4x^4y^4 \cdot 2x \cdot 5y^2$$

and simplify your answer as much as possible.

d. Find the greatest common factor of the expressions $14t^6y$ and $4tu^5y^8$.

Simplify your answer as much as possible.

e. Graph the line with slope -7 passing through the point $(-3, -2)$.

f. Write an equation for the line that passes through the point $(-5, 3)$ and is perpendicular to the line $8x + 5y = 11$.

A plausible *prerequisite diagram* of mastery dependencies for the six Elementary Algebra problems may look like in Figure 1. (Reflexivity and transitivity are assumed to hold and are not explicitly depicted.) The mastery of Problem b is, for instance, a prerequisite for the mastery of Problem e. In other words, the mastery of Problem e implies that of Problem b.

[Figure 1]

The prerequisite diagram in Figure 1 completely specifies the feasible *knowledge states*. A respondent can certainly master just Problem a. This does not imply mastery of any other problem. In that case, the knowledge state is $\{a\}$. However, if the respondent masters e, for instance, then a, b, and c must also be mastered. This gives the knowledge state $\{a, b, c, e\}$. In this way, one obtains exactly 10 knowledge states *consistent* with the

prerequisite diagram:

$$\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{a, b, c\},$$

$$\{a, b, c, d\}, \{a, b, c, e\}, \{a, b, c, d, e\}, \{a, b, c, d, e, f\}\}.$$

This set $\mathcal{K}$ of all possible knowledge states is called *knowledge structure*.

These notions are next formalized mathematically in the following section.

2.2 Basic deterministic concepts

A general concept is that of a *knowledge structure*.

Definition 1 *A knowledge structure is a pair* $(Q, \mathcal{K})$, *with* Q *a non-empty, finite set, and* $\mathcal{K}$ *a family of subsets of* Q *containing at least the empty set* $\emptyset$ *and* Q. *The set* Q *is called the domain of the knowledge structure. The elements* $q \in Q$ *and* $K \in \mathcal{K}$ *are referred to as (test) items and (knowledge) states, respectively. We also say that* $\mathcal{K}$ *is a knowledge structure on* Q.

The set Q is assumed to be a set of dichotomous items. In this article, Q is interpreted as a set of questions/problems that can either *be solved* (coded 1) or *not be solved* (coded 0). This stands for the observed responses of a subject (manifest level), and has to be distinguished from a subject's true, unobservable knowledge of the solution to an item (latent

level). In the latter case, we say that the subject is *capable of mastering* or *not capable of mastering* the item. Let 2^Q denote the *power-set* of Q, that is, the set of all subsets of Q. Let $|Q|$ stand for the *size* (number of elements) of Q. The observed responses of a subject are represented by the subset $R \subset Q$ containing exactly the items solved by the subject. This subset R is called the *response pattern* of the subject. Similarly, the true latent state of knowledge of a subject is represented by the subset $K \subset Q$ containing exactly the items the subject is capable of mastering. This subset K is called the *knowledge state* of the subject. Given a knowledge structure $\mathcal{K}$, the only states of knowledge possible are assumed to be the ones in $\mathcal{K}$. In this spirit, $\mathcal{K}$ captures the organization of knowledge in the domain and population under reference. Idealized, if no response errors would be committed, the only response patterns possible would be the knowledge states in $\mathcal{K}$.

As an example knowledge structure consider the one described in Section 2.1, on the domain $Q = \{a, b, c, d, e, f\}$ of the six Elementary Algebra problems.

Note that this example knowledge structure is *closed under union* and *intersection.*

Definition 2 *A knowledge structure $(Q, \mathcal{K})$ is called a knowledge space if and only if (iff) the union of any two knowledge states is a knowledge state. A knowledge space $(Q, \mathcal{K})$ is called quasi-ordinal iff the intersection of any two knowledge states is a knowledge state.*

The notions of a knowledge structure and (quasi-ordinal) knowledge space are at the level of persons (representing collections of knowledge states of individuals). There is another important notion, that of a *surmise relation*, which is at the level of items (representing collections of mastery dependencies between items).

Definition 3 *Any quasi-order, that is, reflexive and transitive binary relation, on Q is called a surmise relation.*

A surmise relation $\prec$ on Q may model a latent hierarchy among the items based on mastery dependencies of the following type: a subject capable of mastering item $j \in Q$ is also capable of mastering item $i \in Q$ (i.e., $i \prec j$).

As an example surmise relation consider the surmise relation $\prec$ corresponding to the prerequisite diagram of mastery dependencies in Figure 1: $a \prec a$, $b \prec b$, $c \prec c$, $d \prec d$, $e \prec e$, $f \prec f$, $a \prec c$, $a \prec d$, $a \prec e$, $a \prec f$, $b \prec d$, $b \prec e$, $b \prec f$, $c \prec d$, $c \prec e$, $c \prec f$, $d \prec f$, $e \prec f$.

Birkhoff's (1937) theorem (applied in KST, Theorem 1) provides a linkage between quasi-ordinal knowledge spaces and surmise relations on an item set. (This theorem is crucial in this article. It allows for formulating the modified Item Tree Analysis procedure (Section 5.2) and comparing the coefficients κ and CA (Section 6.3).)

Theorem 1 *There is a one-to-one correspondence between the family of all quasi-ordinal knowledge spaces $\mathcal{K}$ on a domain Q, and the family of all surmise relations $\prec$ on Q. Such a correspondence is defined through the two equivalences:*

$$a \prec b \quad \textit{iff} \quad \textit{for any state } K \in \mathcal{K}, \, b \in K \textit{ implies } a \in K;$$

$$K \in \mathcal{K} \quad \textit{iff} \quad \textit{for any pair } a \prec b, \, b \in K \textit{ implies } a \in K.$$

Proof See Doignon and Falmagne (1999, pp. 39-40, Theorem 1.49). □

This theorem is important from a practical point of view. Though the quasi-ordinal knowledge space and surmise relation models are empiri-

cally interpreted at two different levels, at the levels of persons and items respectively, they are connected with each other, through Birkhoff's theorem, on a solid mathematical basis. Roughly speaking, it mathematically links two different levels of empirical interpretations.

In the example in Section 2.1, the 10 knowledge states consistent with the prerequisite diagram in Figure 1 are obtained by applying the second equivalence of Birkhoff's theorem.

2.3 Basic probabilistic concepts

Examinees are drawn randomly from a population under reference. Let N be the sample size. The data are represented by the absolute counts $N(R)$ of response patterns (subsets of Q containing exactly the items solved by the examinees) $R \subset Q$. We assume that the examinees give their response patterns independent of each other. The true probability of occurrence $\rho(R)$ of any response pattern $R \subset Q$ is assumed to stay constant across the examinees. Hence the data are assumed to be *multinomially* distributed over all subsets of Q.

Let the *maximum* probability of occurrence be denoted by $\rho(R_{max})$,

$$\rho(R_{max}) = \max_{R \subset Q} \rho(R),$$

for some appropriate response pattern $R_{max} \subset Q$. *Maximum likelihood estimates* (MLEs) for the population probabilities $\rho(R)$ ($R \subset Q$) are $\widehat{\rho(R)} = N(R)/N$. The MLE for $\rho(R_{max})$ is

$$\widehat{\rho(R_{max})} = N(R'_{max})/N,$$

where $N(R'_{max})$ denotes the *maximum* absolute count $N(R'_{max}) = \max_{R \subset Q} N(R)$, for some appropriate response pattern $R'_{max} \subset Q$.

In the example in Section 6, we simulate multinomial response data using a *basic local independence model.*

Definition 4 *A quadruple $(Q, \mathcal{K}, p, r)$ is called a basic local independence model (BLIM) iff*

1. *$(Q, \mathcal{K})$ is a knowledge structure;*

2. *p is a probability distribution on $\mathcal{K}$, that is, $p(K) > 0$ for any $K \in \mathcal{K}$, and $\sum_{K \in \mathcal{K}} p(K) = 1$;*

3. *r is a response function for $(Q, \mathcal{K}, p)$, that is, $r(R, K) \geq 0$ for any $R \subset Q$ and $K \in \mathcal{K}$, and $\sum_{R \subset Q} r(R, K) = 1$ for any $K \in \mathcal{K}$;*

4. r satisfies local independence, that is, for any $R \subset Q$ and $K \in \mathcal{K}$,[3]

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

with two constants $\beta_q, \eta_q \in [0, 1)$ for each $q \in Q$, respectively called careless error and lucky guess probabilities at q.

To each knowledge state $K \in \mathcal{K}$ is attached a probability $p(K)$ measuring the likelihood that a randomly sampled examinee is in state K (Part 2). For $R \subset Q$ and $K \in \mathcal{K}$, $r(R, K)$ specifies the conditional probability of response pattern R for an examinee in state K (Part 3). (The BLIM takes into account the two ways in which probabilities must supplement deterministic knowledge structures. First, knowledge states may occur with different proportions in the population under reference. Second, response

---

[3] Note that $K \setminus R = \{q \in Q : q \in K$ and $q \notin R\}$, $K \cap R = \{q \in Q : q \in K$ and $q \in R\}$, $R \setminus K = \{q \in Q : q \in R$ and $q \notin K\}$, and $Q \setminus (R \cup K) = \{q \in Q : q \notin R$ and $q \notin K\}$ form a *partition* of Q, that is, $Q = (K \setminus R) + (K \cap R) + (R \setminus K) + (Q \setminus (R \cup K))$. Roughly speaking, items in $K \setminus R$, $K \cap R$, $R \setminus K$, and $Q \setminus (R \cup K)$ are mastered but not solved (careless error), mastered and solved (no careless error), solved but not mastered (lucky guess), and not solved and not mastered (no lucky guess), respectively.

errors (careless errors and lucky guesses) may render impossible a pri-
ori specification of the observable responses of an examinee given her/his
knowledge state.) The item responses of an examinee are assumed to be
independent given the knowledge state of the examinee, and the response
error probabilities β_q, η_q $(q \in Q)$ are attached to the items (item-specific)
and do not vary from state to state (state-independent) (Part 4).

Under the BLIM, the manifest multinomial probability distribution on
the response patterns is governed by the latent state proportions and
response error rates:

$$\rho(R) = \sum_{K \in \mathcal{K}} \left\{ \left[\prod_{q \in K \setminus R} \beta_q \right] \cdot \left[\prod_{q \in K \cap R} (1 - \beta_q) \right] \cdot \left[\prod_{q \in R \setminus K} \eta_q \right] \cdot \left[\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right] \right\} p(K).$$

3 Correlational Agreement Coefficient CA

This section briefly reviews the Correlational Agreement Coefficient CA.
For details, the reader is referred to Ünlü and Albert (2004); see also
Schrepp (2006).

The Correlational Agreement Coefficient CA was introduced by Leeuwe
(1974) within Item Tree Analysis, a data-analytic procedure for deriving

surmise relations on sets of dichotomous items (Section 5.1). In Item Tree Analysis, CA is used as a descriptive goodness-of-fit measure for selecting out of competing surmise relations one with maximum CA value. It is a measure formulated at the level of items, for surmise relations (cf. Theorem 1).

3.1 Required notation and terminology

Let the non-empty, finite item set be denoted by $Q = \{I_l : 1 \leq l \leq m\}$. (The definition of CA requires an indexing of the items. That is why we use this notation.) For the random sample of N examinees, let the corresponding binary (of type 0/1) data matrix (of item responses) be D. Let $\prec$ be a surmise relation on Q. We say that $\prec$ is *consistent* with the data matrix D iff for any item pair $I_i \prec I_j$, every examinee solving item I_j also solves item I_i.

Empirical correlation r_{ij} between items I_i and I_j is defined as the sample Pearson correlation between the corresponding columns s_i and s_j of D. That is, $r_{ij} = Cov(s_i, s_j)/(\sqrt{Var(s_i)}\sqrt{Var(s_j)})$, where $Cov()$ and $Var()$ stand for the sample covariance and variance, respectively.

Theoretical correlation r_{ij}^* between items I_i and I_j is defined as

$$
r_{ij}^* = \begin{cases}
1 & : \quad I_i \prec I_j \text{ and } I_j \prec I_i \\[2mm]
\sqrt{\dfrac{(1-p_{I_i})\cdot p_{I_j}}{(1-p_{I_j})\cdot p_{I_i}}} & : \quad I_i \prec I_j \text{ and } I_j \nprec I_i \\[2mm]
\sqrt{\dfrac{(1-p_{I_j})\cdot p_{I_i}}{(1-p_{I_i})\cdot p_{I_j}}} & : \quad I_i \nprec I_j \text{ and } I_j \prec I_i \\[2mm]
0 & : \quad I_i \nprec I_j \text{ and } I_j \nprec I_i
\end{cases}
$$

where p_{I_i} and p_{I_j} are the sample proportions-correct of items I_i and I_j, respectively.

3.2 Definition of CA

A comparison of empirical and theoretical correlation gives the following result.

Proposition 1 *Let $\prec$ be a surmise relation on Q that is consistent with the data matrix D. Let I_i and I_j be items for which the empirical correlation exists. For the difference $\delta_{ij} = r_{ij} - r_{ij}^*$ between empirical and theoretical correlation, it holds*

$$
\delta_{ij} \begin{cases}
= 0 & : \quad I_i \prec I_j \text{ or } I_j \prec I_i \\[2mm]
\neq 0 \text{ in general} & : \quad I_i \nprec I_j \text{ and } I_j \nprec I_i
\end{cases}
$$

Proof See Ünlü and Albert (2004, p. 287, Proposition 12). □

Proposition 1 gives motivation for the definition of the *Correlational Agreement Coefficient.*

Definition 5 *The Correlational Agreement Coefficient CA is defined by*

$$CA = CA(\prec, D) = 1 - \frac{2}{m(m-1)} \sum_{(I_i, I_j) \in A_Q} (r_{ij} - r_{ij}^*)^2,$$

where

$$A_Q = \{(I_i, I_j) \in Q \times Q : \ i < j \ and \ r_{ij} \ exists\}.$$

The decision rule for applications of CA is as follows. The greater the value of CA is, the 'better' a surmise relation is judged to fit the data.

4 Measure κ

This section proposes the measure κ for evaluating knowledge structures. It is specially designed to trade off the fit and size of a knowledge structure, and is derived within an operational framework (prediction paradigm).

4.1 Prediction paradigm

The *prediction problem* considered is as follows. An individual is randomly chosen from the population under reference, and we are asked to guess her/his response pattern, given either

(no info). no further information (other than the multinomial distribu-

   tion), or

(info). the knowledge structure $\mathcal{K}$ assumed to underlie the responses of

   the individual.

The *prediction strategies* in the two cases are as follows. In the 'no info'
case, we *optimally* guess some response pattern $R_{max} \subset Q$ with the largest
probability of occurrence $\rho(R_{max})$. In the 'info' case, we *proportionally*
guess the knowledge states $K \in \mathcal{K}$ with their probabilities of occurrence
$\rho(K)$. Since the latter probabilities may not add up to one, in general,
there may be a non-vanishing *residual* probability $1 - \sum_{K \in \mathcal{K}} \rho(K) > 0$.
To complete the prediction strategy, hence we abstain from guessing with
probability $1 - \sum_{K \in \mathcal{K}} \rho(K)$, and in the sequel, view that as a prediction
error.

The *probabilities of a prediction error* in the two cases are as follows. In
the 'no info' case, the probability of a prediction error is $1 - \rho(R_{max})$; in
the 'info' case, it is $1 - \sum_{K \in \mathcal{K}} \rho^2(K)$. (The (complementary) *probabilities
of a prediction success* are $\rho(R_{max})$ and $\sum_{K \in \mathcal{K}} \rho^2(K)$, respectively.)

4.2 First constituent of κ: Measure of fit

The measure κ consists of two constituents. The first constituent of κ measures the degree to which a knowledge structure *descriptively* reflects the response data; the *fit*. It expresses the extent to which the multinomial probability distribution on the response patterns is concentrated to the knowledge structure.

The first constituent of κ is derived on the basis of the method of *proportional reduction in predictive error* (PRPE); the method of PRPE was introduced by Guttman (1941), and it was systematically applied in the series of articles by Goodman and Kruskal (1954, 1959, 1963, 1972). The general probability formula of the method of PRPE quantifies the *predictive utility*, PU_{info}, of given information:

$$PU_{info} = \frac{\text{Prob. of error (no info)} - \text{Prob. of error (info)}}{\text{Prob. of error (no info)}}.$$

Inserting the afore mentioned prediction error probabilities into this formula, we obtain the population analog of the first constituent m_1,

$$m_1 = \frac{\sum_{K \in \mathcal{K}} \rho(K)^2 - \rho(R_{max})}{1 - \rho(R_{max})}.$$

Some remarks are in order with respect to m_1.

1. It holds $-\infty < m_1 \le 0$ ($\rho(R_{max}) \ge \sum_{R \subset Q} \rho^2(R) \ge \sum_{K \in \mathcal{K}} \rho^2(K)$).

2. Obviously, $m_1 = 0$ iff $\rho(R_{max}) = \sum_{K \in \mathcal{K}} \rho^2(K)$. In other words, m_1 assumes its extreme value in the case of, and only of, guessing with the largest probability of a prediction success. In that case, we have zero residual probability ($\sum_{K \in \mathcal{K}} \rho(K) = 1$), and the distribution on the response patterns is completely concentrated to the knowledge structure $\mathcal{K}$ ($\rho(R) = 0$ for any $R \subset Q$, $R \notin \mathcal{K}$).

Inserting MLEs, we obtain the MLE $\widehat{m_1}$ for m_1,

$$\widehat{m_1} = \frac{\sum_{K \in \mathcal{K}} N(K)^2 - N \cdot N(R'_{max})}{N^2 - N \cdot N(R'_{max})}.$$

(We assume that $1 - N(R'_{max})/N \ne 0$. Since, by assumption, $\rho(R_{max}) \ne 1$, and $N(R'_{max})/N$ is the MLE for $\rho(R_{max})$, this is likely the case for large samples.)

4.3 Second constituent of κ: Measure of size

The second constituent of κ captures the *size* of a knowledge structure. It expresses the extent to which the restricted multinomial probability distribution on the knowledge states is concentrated to a fraction of the knowledge structure. (In Section 4.5, a special choice of a fraction is de-

termined in the context of 'model selection' among competing knowledge structures, based on the *median match* of the competing models.)

The definition of the second constituent of κ is based on the following notion of a *truncation* of a knowledge structure. Let $n \geq 1$ (a natural number) be a *truncation constant*. An *n-truncation* of $\mathcal{K}$ is any subset $\mathcal{K}_{nt}$ of $\mathcal{K}$ which is derived as follows:

1. Order the knowledge states according to their occurrence probabilities, say, from left to right, ascending from smaller probabilities to larger ones. Knowledge states with equal probabilities are ordered arbitrarily.

2. Starting with the foremost right knowledge state, a knowledge state with the largest probability of occurrence, take the first $\min(|\mathcal{K}|, n)$ knowledge states, descending from right to left. The set of all these knowledge states is called an *n*-truncation of $\mathcal{K}$, denoted by $\mathcal{K}_{nt}$.

(Depending on the orderings of equiprobable knowledge states, the set $\mathcal{K}_{nt}$ may vary. In general, there are multiple *n*-truncations of a knowledge structure. The definition of the second constituent, however, is invariant with respect to the choice of a particular *n*-truncation. In Section 4.5, we

describe how a reasonable truncation constant can be chosen in the context of 'model selection' among competing knowledge structure models.)

For a truncation constant n, and any n-truncation $\mathcal{K}_{nt}$ of $\mathcal{K}$, we obtain the population analog of the second constituent m_2,

$$m_2 = \frac{\sum_{K \in \mathcal{K}} \rho^2(K)}{\sum_{K \in \mathcal{K}_{nt}} \rho^2(K)}.$$

Some remarks are in order with respect to m_2.

1. It holds $1 \leq m_2 < +\infty$.

2. Obviously, $m_2 = 1$ iff $\sum_{K \in \mathcal{K}_{nt}} \rho^2(K) = \sum_{K \in \mathcal{K}} \rho^2(K)$. In other words, m_2 assumes its extreme value in the case of, and only of, no loss of the probability of a prediction success when guessing based on a fraction, an n-truncation, of $\mathcal{K}$ than on the whole knowledge structure $\mathcal{K}$. In that case, $\sum_{K \in \mathcal{K}_{nt}} \rho(K) = \sum_{K \in \mathcal{K}} \rho(K)$, and the restricted distribution on the knowledge states is completely concentrated to an n-truncation $\mathcal{K}_{nt}$ ($\rho(K) = 0$ for any $K \in \mathcal{K}$, $K \notin \mathcal{K}_{nt}$).

Inserting MLEs, we obtain the MLE $\widehat{m_2}$ for m_2,

$$\widehat{m_2} = \frac{\sum_{K \in \mathcal{K}} N(K)^2}{\sum_{K \in \widehat{\mathcal{K}_{nt}}} N(K)^2},$$

where $\widehat{\mathcal{K}_{nt}}$ is analogously defined as $\mathcal{K}_{nt}$, replacing occurrence probabilities $\rho(K)$ with their MLEs $N(K)/N$ (for knowledge states $K \in \mathcal{K}$). (We assume that $\sum_{K \in \mathcal{K}} N(K) \neq 0$. Since, by assumption, $\sum_{K \in \mathcal{K}} \rho(K) \neq 0$, and $(\sum_{K \in \mathcal{K}} N(K))/N$ is the MLE for $\sum_{K \in \mathcal{K}} \rho(K)$, this is likely the case for large samples.)

4.4 Measure κ: Size/fit trade-off

The measure κ is defined as the product of the size measure m_2 and the shifted fit measure m_1. (The shift by '$-c$' of the fit measure compensates for a zero value of that measure. This is necessary to guarantee a trade-off between the size and fit criteria. For more details, see below.)

Definition 6 *Let $n \geq 1$ be a truncation constant, and let $c \in [0, 0.01]$ be a non-negative shift constant. The measure κ is defined by*

$$\kappa = m_2(m_1 - c) = \frac{\sum_{K \in \mathcal{K}} \rho^2(K)}{\sum_{K \in \mathcal{K}_{nt}} \rho^2(K)} \left(\frac{\sum_{K \in \mathcal{K}} \rho^2(K) - \rho(R_{max})}{1 - \rho(R_{max})} - c \right).$$

Some remarks are in order with respect to κ.

1. It holds $-\infty < \kappa \leq -c$.

2. For $c = 0$, $\kappa = -c(= 0)$ iff $m_1 = 0$, with in general arbitrary values of m_2. In other words, κ assumes its extreme value in the case of, and

only of, 'complete association' as described for the fit measure m_1, and there is no indication about the size component as measured by the size measure m_2. In that case, $\sum_{K \in \mathcal{K}} \rho(K) = 1$, and there is a total fit of $\mathcal{K}$ to the data. (If $\kappa = m_2 m_1$, that is, $c = 0$, a zero value of the fit measure m_1 would eliminate the impact of the size measure m_2 on the values of κ. Regardless of any value m_2 may take, κ would always be equal to zero. In that case, κ would not trade off the size and fit components.)

3. For $c \neq 0$, $\kappa = -c(< 0)$ iff $m_1 = 0$ and $m_2 = 1$. In other words, κ assumes its extreme value in the case of, and only of, 'complete associations' as described for the measures m_1 and m_2. In that case, (a) $\sum_{K \in \mathcal{K}} \rho(K) = 1$, and there is a total fit of $\mathcal{K}$ to the data, and (b) $\rho(K) = 0$ for any $K \in \mathcal{K}$, $K \notin \mathcal{K}_{nt}$, and the size of $\mathcal{K}$ can actually be reduced to $|\mathcal{K}_{nt}|$.

4. Larger values of κ imply larger values of m_1, or smaller values of m_2, or both. That is, if $\kappa(\mathcal{K}_1)$ and $\kappa(\mathcal{K}_2)$ denote the measure κ calculated for the knowledge structures $\mathcal{K}_1$ and $\mathcal{K}_2$, respectively, $\kappa(\mathcal{K}_1) < \kappa(\mathcal{K}_2)$ implies $m_1(\mathcal{K}_1) < m_1(\mathcal{K}_2)$ or $m_2(\mathcal{K}_1) > m_2(\mathcal{K}_2)$. The reverse needs

not to be true in general. However, $m_1(\mathcal{K}_1) < m_1(\mathcal{K}_2)$ and $m_2(\mathcal{K}_1) > m_2(\mathcal{K}_2)$ implies $\kappa(\mathcal{K}_1) < \kappa(\mathcal{K}_2)$.

5. The afore mentioned remark can be rephrased in operational parlance of the prediction paradigm. Larger values of κ imply larger probabilities of a prediction success, or larger relative probabilities of a prediction success for n-truncations, or both. In that case, $\mathcal{K}_2$ 'performs better' than $\mathcal{K}_1$ with respect to at least one of the criteria size and fit. If $\mathcal{K}_2$ 'performs better' than $\mathcal{K}_1$ with respect to both of the criteria, it necessarily holds $\kappa(\mathcal{K}_1) < \kappa(\mathcal{K}_2)$. If $\mathcal{K}_2$ 'performs better' than $\mathcal{K}_1$ with respect to one of the two criteria only, $\kappa(\mathcal{K}_1) < \kappa(\mathcal{K}_2)$ may not be true in general.

Inserting the MLEs $\widehat{m_1}$ and $\widehat{m_2}$ for m_1 and m_2, respectively, we obtain the MLE $\widehat{\kappa}$ for κ,

$$\widehat{\kappa} = \widehat{m_2}(\widehat{m_1} - c) = \frac{\sum_{K \in \mathcal{K}} N(K)^2}{\sum_{K \in \widehat{\mathcal{K}_{nt}}} N(K)^2} \left(\frac{\sum_{K \in \mathcal{K}} N(K)^2 - N \cdot N(R'_{max})}{N^2 - N \cdot N(R'_{max})} - c \right).$$

The decision rule for applications of κ is as follows. The greater the (population) value of κ is, the 'better' a knowledge structure 'performs' with respect to a trade-off between the criteria size and fit. The unknown

ordering of the (population) κ values is estimated by the ordering of the corresponding MLEs.

4.5 *Model selection and truncation constant*

Next we describe a special choice for the truncation constant. That special truncation constant is derived in the context of 'model selection' among competing knowledge structures, say, $\mathcal{K}_1, \ldots, \mathcal{K}_p$ (on a domain Q).

The definition of the special truncation constant is based on the notion of the *median match* of the competing models under consideration. Let $v_i = |\{K \in \mathcal{K}_i : \rho(K) \neq 0\}|$ be the *match* of the candidate model $\mathcal{K}_i$ $(1 \leq i \leq p)$. Let $\mathbf{v} = (v_1, \ldots, v_p)$ be the *match vector* of the 'model selection' problem. The (empirical) median of the matches v_i is denoted by $median(\mathbf{v})$, and is called the *median match* of the competing models of the 'model selection' problem. That is,

$$
median(\mathbf{v}) = \begin{cases} v_{(\frac{p+1}{2})} & : \quad \text{odd } p \\[2ex] v_{(\frac{p}{2})} & : \quad \text{even } p \end{cases}
$$

where $v_{(1)}, \ldots, v_{(p)}$ with $v_{(1)} \leq \cdots \leq v_{(p)}$ is the ordered list of the matches v_i.

For efficient applications of adaptive knowledge assessment procedures (Doignon and Falmagne, 1999), knowledge structures of a 'trade-off type' are beneficial. On the one hand, a knowledge structure should fit the data as well as possible, but on the other hand, it should also be of a smaller size. A trade-off between these criteria allows for an economic diagnosis of the knowledge state of an examinee; in general, for a knowledge structure of a smaller size only a fewer items have to be answered by an examinee to assess her/his knowledge state. The *half-split rule* in deterministic knowledge assessment, for instance, generally requires about $\log_2(|\mathcal{K}|)$ items for the diagnosis. (At this point it should be clear why a knowledge structure is not preferred to consist of all subsets of an item set. In such a case, no dependencies between the items would be postulated (except for reflexive ones), and hence all items of the domain would have to be worked through by an examinee. Compare also the simulation example in Section 6.) In other words, the measure κ (as a size/fit trade-off procedure) allows for a 'poorer' (descriptive) fit of a knowledge structure in favor of a smaller size.

The definition of the special truncation constant is furthermore based on the term $2^{|Q|/2}$. This term is introduced to express the extent to which a knowledge structure may be 'tailored' to n-truncations of sizes bounded from above by $2^{|Q|/2}$. (The half-split rule using a knowledge structure of a size bounded from above by $2^{|Q|/2}$, generally requires about at most half of the items for the diagnosis of the knowledge state of an examinee $(\log_2(|\mathcal{K}|) \le \log_2(2^{|Q|/2}) = |Q|/2).)$

The *special truncation constant n_s* is defined by

$$n_s = \min\left([2^{|Q|/2}], median(\mathbf{v})\right),$$

where for a real number $x \ge 0$, $[x]$ denotes the *entier* of x, that is, the non-negative integer k with $k \le x < k + 1$.

The MLE for the (population) special truncation constant is

$$\widehat{n}_s = \min\left([2^{|Q|/2}], median(\widehat{\mathbf{v}})\right),$$

where $\widehat{v}_i = |\{K \in \mathcal{K}_i : N(K) \ne 0\}|$ $(1 \le i \le p)$ and $\widehat{\mathbf{v}} = (\widehat{v}_1, \ldots, \widehat{v}_p)$ are the MLEs for the (population) matches and match vector, respectively.

4.6 Summary

In the example in Section 6, the measure κ is applied using the special truncation constant. Next we summarize the (population) κ procedure based on the special truncation constant, and the corresponding MLE in terms of the data.

Consider the 'model selection' problem

$$\mathcal{K}_1, \ldots, \mathcal{K}_p$$

with knowledge structures $\mathcal{K}_i$ on Q. Assume that for the multinomial probability distribution on the response patterns, it holds

1. $\rho(R_{max}) \neq 1$ $(R_{max} \subset Q$ with $\rho(R_{max}) = \max_{R \subset Q} \rho(R))$;

2. $\sum_{K \in \mathcal{K}_i} \rho(K) \neq 0$ for any $\mathcal{K}_i$.

Let n_s be the (population) special truncation constant, and let $c \in [0, 0.01]$ be a shift constant.

Under these conditions the (population) $\kappa(\mathcal{K}_i)$ values for the knowledge structures $\mathcal{K}_i$ are well-defined,

$$\kappa(\mathcal{K}_i) = \frac{\sum_{K \in \mathcal{K}_i} \rho^2(K)}{\sum_{K \in \mathcal{K}_{inst}} \rho^2(K)} \left(\frac{\sum_{K \in \mathcal{K}_i} \rho^2(K) - \rho(R_{max})}{1 - \rho(R_{max})} - c \right).$$

Let $N(R)$ for $R \subset Q$ be the data, and let $N = \sum_{R \subset Q} N(R)$ be large.

Assume that

1. $N(R'_{max}) \neq N$ ($R'_{max} \subset Q$ with $N(R'_{max}) = \max_{R \subset Q} N(R)$);

2. $\sum_{K \in \mathcal{K}_i} N(K) \neq 0$ for any $\mathcal{K}_i$.

Inserting (sample) MLEs

$$N(K)/N \text{ for } \rho(K) \quad (K \in \mathcal{K}_i),$$

we obtain the (sample) MLE $\widehat{\kappa(\mathcal{K}_i)}$ for $\kappa(\mathcal{K}_i)$,

$$\widehat{\kappa(\mathcal{K}_i)} = \frac{\sum_{K \in \mathcal{K}_i} N(K)^2}{\sum_{K \in \widehat{\mathcal{K}_{i\widehat{n}_s t}}} N(K)^2} \left(\frac{\sum_{K \in \mathcal{K}_i} N(K)^2 - N \cdot N(R'_{max})}{N^2 - N \cdot N(R'_{max})} - c \right),$$

where $\widehat{n}_s$ is the (sample) MLE for n_s, and $\widehat{\mathcal{K}_{i\widehat{n}_s t}}$ is analogously defined as $\mathcal{K}_{i\widehat{n}_s t}$, replacing occurrence probabilities $\rho(K)$ with their MLEs $N(K)/N$ (for knowledge states $K \in \mathcal{K}_i$).

Note that *all* $\mathcal{K}_i$ ($1 \leq i \leq p$) are used for determining the (population) n_s and (sample) $\widehat{n}_s$ values, and based on these n_s and $\widehat{n}_s$ values, the (population) $\kappa(\mathcal{K}_i)$ and (sample) $\widehat{\kappa(\mathcal{K}_i)}$ values are obtained for each model *separately*, respectively.

Eventually, a model $\mathcal{K}_{i_0}$ $(1 \leq i_0 \leq p)$ is selected for which $\widehat{\kappa(\mathcal{K}_{i_0})}$ is maximum, that is,

$$\widehat{\kappa(\mathcal{K}_{i_0})} = \max_{1 \leq i \leq p} \widehat{\kappa(\mathcal{K}_i)}.$$

5 Modified Item Tree Analysis

This section describes how the candidate competing models are obtained data-analytically, based on a modified version of Leeuwe's (1974) Item Tree Analysis. (Though we pursue a data-analytic approach, any other method for determining the competing models is conceivable. The measure κ can be applied to any 'model selection' problem among competing knowledge structures, independent of how the models may be obtained. For instance, κ may be used for selecting among knowledge structures theoretically derived from different psychological theories/postulates.)

5.1 Item Tree Analysis

This section reviews Leeuwe's (1974) *Item Tree Analysis* (ITA); see also Ünlü and Albert (2004).

ITA consists of five steps, STEP1–STEP5:

STEP1. Determine the binary relations $\prec_L$ for $L = 0, 1, \ldots, N$ according to the *ITA-rule*

$$I_i \prec_L I_j \quad \text{iff} \quad c_{ij} \leq L.$$

Here we use the notation $Q = \{I_l : 1 \leq l \leq m\}$, and for any two items I_i and I_j in Q, c_{ij} denotes the absolute count of examinees solving I_j but not I_i. The *tolerance level L* quantifies the allowed maximum number of contradictions to an item pair in the relation $\prec_L$.

STEP2. From the generated binary relations $\prec_L$ ($0 \leq L \leq N$), remove those that are not transitive.

STEP3. Set a critical value $0 < c \leq 1$ for the proportions p_L of examinees not contradicting the respective *surmise relations* $\prec_L$ in STEP2.

STEP4. From the surmise relations in STEP2, remove those with $p_L < c$.

STEP5. From the remaining surmise relations (after STEP4)—$\prec_0$ is always contained—select one with maximum CA value.

(The Correlational Agreement Coefficient is used as a goodness-of-fit measure to handle the selection problem in STEP5. From the remaining surmise relations, select an 'optimal' one, here, one with maximum CA value.)

5.2 Modified version of Item Tree Analysis

A *modified version of ITA* (MITA) is as follows. We keep the ITA-rule for generating the binary relations $\prec'_L$ for $0 \leq L \leq N$. However, we do not remove those relations $\prec'_L$ that are not transitive, as it is done in the original ITA procedure. Instead, we take the *transitive closure* of such a binary relation turning it to a surmise relation. Moreover, the other steps of ITA are not considered anymore. This yields a collection of candidate models which contains the models derived from ITA. Eventually, from the collection of candidate surmise relation models, we select an 'optimal' one, here, one with maximum κ value. (Note that Theorem 1 is crucial at this point. We select among the quasi-ordinal knowledge spaces corresponding to the surmise relations according to Theorem 1. The measure κ is formulated at the level of persons, for knowledge structures.)

This is the *fully-automated* version of MITA. In a *user-controlled* version, the user may narrow down this collection of competing models to a smaller one, based on important factors (e.g., psychological theory) not captured by the data analysis solely. (The fully-automated version of MITA is illustrated in Section 6.)

MITA consists of three steps, STEP1–STEP3:

STEP1. Determine the binary relations $\prec'_L$ for $L = 0, 1, \ldots, N$ according to the ITA-rule.

STEP2. Take the transitive closure for all $\prec'_L$ ($0 \leq L \leq N$). Consider the collection of (quasi-ordinal) knowledge spaces corresponding to those surmise relations according to Theorem 1. In the fully-automated version of MITA, these knowledge spaces constitute the final collection of competing models. In a user-controlled version, that collection is further narrowed down to a smaller sub-collection, based on important factors not captured by the data analysis solely.

STEP3. From the collection of knowledge structure models in STEP2, select one with maximum κ value.

6 Simulation example

This section describes an application of the coefficients κ and CA. Their performances as selection measures are compared in a simulation study using the BLIM (Definition 4). (The computer programs (in `C`) for the computations in this section can be obtained from the first author.)

6.1 Data generating model

We consider the knowledge structure

$$\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, \{a, b, d\}, \{a, b, c, d\}, \{a, b, c, e\}, Q\}$$

on the domain $Q = \{a, b, c, d, e\}$. The *Hasse diagram* of the surmise relation $\prec_{\mathcal{K}}$ derived from $\mathcal{K}$ according to Theorem 1 is shown in Figure 2.

[Figure 2]

We assume that the knowledge states of $\mathcal{K}$ occur in the population under reference with the probabilities

$$p(\emptyset) = 0.04,$$

$$p(\{a\}) = 0.10,$$

$$p(\{b\}) = 0.06,$$

$$p(\{a, b\}) = 0.12,$$

$$p(\{a, b, c\}) = 0.11,$$

$$p(\{a, b, d\}) = 0.07,$$

$$p(\{a, b, c, d\}) = 0.13,$$

$$p(\{a, b, c, e\}) = 0.18,$$

$$p(Q) = 0.19.$$

Let the careless error and lucky guess probabilities β_q and η_q at the items $q \in Q$, respectively, be specified as

$$\beta_a = 0.16, \ \eta_a = 0.04,$$

$$\beta_b = 0.18, \ \eta_b = 0.10,$$

$$\beta_c = 0.20, \ \eta_c = 0.01,$$

$$\beta_d = 0.14, \ \eta_d = 0.02,$$

$$\beta_e = 0.24, \ \eta_e = 0.05.$$

This BLIM was used for the simulation of the data (Section 6.2). Note that the specification of the model parameters $p(K)$ $(K \in \mathcal{K})$ and β_q, η_q $(q \in Q)$ is a realistic one. We do not assume a single response error rate over all items; the careless error and lucky guess rates vary from item to item. We do not assume a uniform probability distribution on $\mathcal{K}$; the knowledge states occur with different proportions in the population under reference. Furthermore, from an empirical point of view, the lower values for the lucky guess rates do not cause concern, because guessing effects can nearly be eliminated by appropriate item formulation.

6.2 Simulated data

We simulated a binary (of type 0/1) $1\,200 \times 5$ data matrix of response patterns for $1\,200$ fictitious subjects. The data matrix contains all of the 32 possible response patterns, hence $1\,168$ response patterns are duplicates. This matrix of item scores is displayed in Table 1. (The 32 response patterns are shown with their absolute frequencies in the data. There are 73 response patterns '00000' (no item solved) and 88 response patterns '11111' (all items solved).)

[Table 1]

6.3 Results: MITA of simulated data

We describe the results of an application of MITA to the simulated BLIM data.

We determined the binary relations $\prec'_L$ ($0 \leq L \leq N = 1\,200$) according to the ITA-rule, and their transitive closures $\prec_L$ ($0 \leq L \leq 1\,200$). This resulted in a collection of 15 surmise relations, respectively, a collection of 15 (quasi-ordinal) knowledge spaces. These collections contained the true models underlying the data, $\prec_{\mathcal{K}}$ and $\mathcal{K}$, respectively. (A complete list of

the competing surmise relation and knowledge structure models in this example can be obtained from the first author.) From these collections, we selected 'optimal' models with maximum CA and κ values. Table 2 reports the values of CA and κ. (For n_s and $c = 0.01$. Models are labeled with their tolerance levels. The true model is indicated by '(true)'. L_{CA} and L_κ denote maximum CA and κ solutions, respectively. A notation '0-58' means that the same surmise relation and knowledge space were obtained for the tolerance levels $0 \leq L \leq 58$.)

[Table 2]

The coefficient CA decreases steadily (except for $L = 96\text{-}100$) with its maximum value assumed at the lowest tolerance range $L_{CA} = 0\text{-}58$. The 'optimal' surmise relation selected based on CA is $\prec_{0\text{-}58}$, which is the *diagonal* $i \prec_{0\text{-}58} i$ $(i \in Q)$. It consists of 5 item pairs of altogether $5^2 = 25$ possible pairs (20%). The knowledge space $\mathcal{K}_{0\text{-}58}$ is the set of all subsets of Q, consisting of 32 knowledge states $(\log_2(|\mathcal{K}_{0\text{-}58}|) = 5)$. These 'best' solutions based on CA do not reflect the true models at all. The true surmise relation $\prec_\mathcal{K}$ $(= \prec_{101\text{-}150})$ and knowledge space $\mathcal{K}$ $(= \mathcal{K}_{101\text{-}150})$ consist of 12 item pairs and 9 knowledge states, respectively.

The coefficient κ assumes its maximum value at the tolerance range $L_\kappa = 192\text{-}213$. The 'optimal' knowledge space selected based on κ is

$$\mathcal{K}_{192\text{-}213} = \{\emptyset, \{a\}, \{a, b\}, \{a, b, c\}, \{a, b, c, d\}, \{a, b, c, e\}, Q\}.$$

It consists of 7 knowledge states ($\log_2(|\mathcal{K}_{192\text{-}213}|) \approx 2.81$). The surmise relation $\prec_{192\text{-}213}$ is 'appropriately sized' (14 item pairs; 56% of all possible pairs). Its Hasse diagram is shown in Figure 3.

[Figure 3]

Compared to the true models underlying the simulation, these 'best' solutions based on κ are quite acceptable. In $\prec_{\mathcal{K}}$, the items a and b, and c and d are not comparable, whereas in $\prec_{192\text{-}213}$, we have $a \prec_{192\text{-}213} b$ and $c \prec_{192\text{-}213} d$. In all other respects, the two surmise relations are identical. We have $|\prec_{\mathcal{K}}| = 12$ versus $|\prec_{192\text{-}213}| = 14$ ($\prec_{\mathcal{K}} \subset \prec_{192\text{-}213}$). In $\mathcal{K}$, the subsets $\{b\}$ and $\{a, b, d\}$ are knowledge states, whereas in $\mathcal{K}_{192\text{-}213}$, they are not. In all other respects, the two knowledge spaces are identical. We have $|\mathcal{K}| = 9$ versus $|\mathcal{K}_{192\text{-}213}| = 7$ ($\mathcal{K}_{192\text{-}213} \subset \mathcal{K}$).

7 Discussion

7.1 Summary

In this article, we have proposed a 'Goodman-Kruskal type' measure κ for selecting among competing knowledge structure models in KST, as underlying latent explanations for discrete multivariate response data. We have utilized this measure in a new 'ITA type' data-analytic method for detecting knowledge structures from data.

This measure κ is suited for nominal data and is (operationally) interpretable in terms of prediction error (success) probabilities of a prediction paradigm. It is designed to combine and trade off the (descriptive) fit and size of a knowledge structure, which is of high interest in KST, especially in the context of adaptive knowledge assessment procedures.

We have compared κ with the Correlational Agreement Coefficient CA, which has been recently discussed as a selection measure for competing surmise relation models. The performances of the two coefficients have been investigated in a simulation study using the fundamental BLIM in KST. Based on the proposed MITA method, the candidate competing surmise relation and knowledge structure models have been obtained. The

'optimal' solutions based on CA have not reflected the true models at all, whereas the solutions based on κ have been quite acceptable.

7.2 Further research

The current simulation study is a starting point for more in-depth analyses of the measure κ. Further research may address in systematic, extensive simulation studies the effects of the variation of the sample size (especially for small sample sizes), the underlying knowledge structure model, and the BLIM parameters. In particular, inferential (asymptotic) statistics (e.g., confidence intervals) and applications to real psychological test data are important and indispensable directions for future research.

Measures of a 'Goodman-Kruskal type' could also be derived within the order-theoretic (at the level of items) formulation of KST, for surmise relations or even surmise systems. The relationship between the set-theoretic and order-theoretic measures could then be investigated. In particular, alternative data analysis methods of an 'ITA type' could be derived, and these procedures could be compared with each other, and especially with other available related or unrelated data-analytic methods.

References

Birkhoff, G. (1937). Rings of sets. *Duke Mathematical Journal* **3**, 443-454.

Doignon, J.-P. and Falmagne, J.-Cl. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies* **23**, 175-196.

Doignon, J.-P. and Falmagne, J.-Cl. (1999). *Knowledge Spaces.* Springer, Berlin.

Falmagne, J.-Cl., Doignon, J.-P., Cosyn, E. and Thiéry, N. (2003). The assessment of knowledge, in theory and in practice. Institute for Mathematical Behavioral Sciences, University of California, Irvine. Paper 26.

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732-764.

Goodman, L. A. and Kruskal, W. H. (1959). Measures of association for cross classifications: II. Further discussion and references. *Journal of the American Statistical Association* **54**, 123-163.

Goodman, L. A. and Kruskal, W. H. (1963). Measures of association for cross classifications: III. Approximate sampling theory. *Journal of the American Statistical Association* **58**, 310-364.

Goodman, L. A. and Kruskal, W. H. (1972). Measures of association for cross classifications: IV. Simplification of asymptotic variances. *Journal of the American Statistical Association* **67**, 415-421.

Guttman, L. (1941). An outline of the statistical theory of prediction. In *The Prediction of Personal Adjustment* (Edited by P. Horst *et al.*), 253-318. Social Science Research Council, New York.

Leeuwe, J. F. J. V. (1974). Item tree analysis. *Nederlands Tijdschrift voor de Psychologie* **29**, 475-484.

Schrepp, M. (2006). Properties of the correlational agreement coefficient: A comment to Ünlü and Albert (2004). *Mathematical Social Sciences* **51**, 117-123.

Ünlü, A. and Albert, D. (2004). The Correlational Agreement Coefficient $CA(\leq, D)$—a mathematical analysis of a descriptive goodness-of-fit measure. *Mathematical Social Sciences* **48**, 281-314.

Figure captions

Figure 1: Prerequisite diagram of mastery dependencies for the six Elementary Algebra problems

Figure 2: Hasse diagram of the true model $(Q, \prec_{\mathcal{K}})$

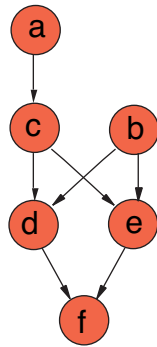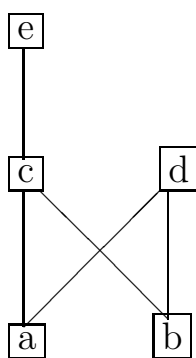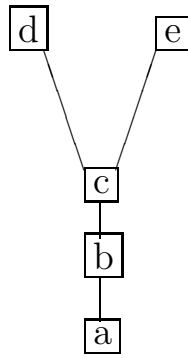Figure 3: Hasse diagram of the data-analytic solution $(Q, \prec_{192\text{-}213})$

Figures

Fig. 1

Fig. 2

Fig. 3

Tables

Table 1 Simulated data for 1 200 fictitious subjects

Pattern	Freq	Pattern	Freq	Pattern	Freq	Pattern	Freq
11001	32	11100	109	01001	13	00001	4
00000	73	11111	88	11000	142	10000	120
11101	110	01000	90	10001	9	01101	16
11110	89	11010	70	00010	5	01100	24
00100	6	01110	16	11011	28	10100	20
01011	4	00101	6	10101	24	10111	21
10010	22	00111	2	01111	13	10110	16
01010	16	00110	6	00011	1	10011	5

A. Ünlü, W. A. Malik

Table 2 Values of CA and κ

L	CA	κ
$L_{CA} = 0\text{-}58$	0.962466	-0.098487
59-62	0.958352	-0.098591
63-71	0.953836	-0.098672
72-77	0.947492	-0.098807
78-88	0.940130	-0.098880
89-95	0.932003	-0.098931
96-100	0.933715	-0.099029
101-150 (true)	0.924552	-0.099040
151-191	0.900744	-0.098871
$L_\kappa = 192\text{-}213$	0.833439	-0.097610
214-236	0.768970	-0.098913
237-239	0.752954	-0.102439
240-285	0.744859	-0.108678
286-394	0.684268	-0.118036
395-1 200	0.319707	-0.133919