

Parameter Estimation of KST-IRT Model under Local Dependence

Sangbeak Ye *, Augustin Kelava and Stefano Noventa

Methods Center, University of Tübingen, 72074 Tübingen, Germany

* Correspondence: sangbeak.ye@uni-tuebingen.de

Abstract: A mantra often repeated in the introductory material to psychometrics and Item Response Theory (IRT) is that a Rasch model is a probabilistic version of a Guttman scale. The idea comes from the observation that a sigmoidal item response function provides a probabilistic version of the characteristic function that models an item response in the Guttman scale. It appears, however, more difficult to reconcile the assumption of local independence, which traditionally accompanies the Rasch model, with the item dependence existing in a Guttman scale. In recent work, an alternative probabilistic version of a Guttman scale was proposed, combining Knowledge Space Theory (KST) with IRT modeling, here referred to as KST-IRT. The present work has, therefore, a two-fold aim. Firstly, the estimation of the parameters involved in KST-IRT models is discussed. More in detail, two estimation methods based on the Expectation Maximization (EM) procedure are suggested, i.e., Marginal Maximum Likelihood (MML) and Gibbs sampling, and are compared on the basis of simulation studies. Secondly, for a Guttman scale, the estimates of the KST-IRT models are compared with those of the traditional combination of the Rasch model plus local independence under the interchange of the data generation processes. Results show that the KST-IRT approach might be more effective in capturing local dependence as it appears to be more robust under misspecification of the data generation process, but it comes with the price of an increased number of parameters.

Keywords: Rasch model; Guttman scale; KST-IRT model; MML; EM algorithm; local dependence



Citation: Ye, S.; Kelava, A.; Noventa, S. Parameter Estimation of KST-IRT Model under Local Dependence. *Psych* **2023**, *5*, 908–927. <https://doi.org/10.3390/psych5030060>

Academic Editor: Okan Bulut

Received: 17 May 2023

Revised: 15 August 2023

Accepted: 18 August 2023

Published: 22 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Local independence (LI) is a property that serves as a fundamental assumption underlying both parametric [1–3] and non-parametric [4–6] IRT models. Despite its importance, LI is easily violated in many situations (see, e.g., [7]). Model-wise, one can generally account for two main types of local dependence (LD), one related to misspecifications in the dimensionality of the model and one related to direct and invasive relations between the items in a test (see, e.g., [8]). In the present work, we are interested in modeling the latter by means of the machinery introduced by KST (see, e.g., [9]). More in detail, we are interested in comparing the modeling of a Guttman scale [10] under two different sets of assumptions: on the one hand, the traditional IRT approach of Rasch plus LI; on the other hand, a latent trait-based version of KST models in which the LD between the items is captured by the existence of a partial order (i.e., a knowledge structure in the shape of a chain of items) of latent states. The assumption of LI is indeed often taken for granted, or its violations are ignored, in the IRT framework for various purposes, including parameter estimation and its use in adaptive testing, and consequently, it can undermine the validity of the estimation (see, e.g., [11]). Yet, the deterministic dependence between responses is a prominent feature of the Guttman scale, which is however difficult to fit into real-world data due to its rigid nature, and thus provides a perfect case study.

As it is often stated in the psychometrics literature, to make the Guttman scale more practical, the work of the early proponents, e.g., Rasch [2] and Mokken [4], evolved the Guttman scale into parametric and non-parametric probabilistic scales, respectively, which, based on the notions of invariant measurement, are generally considered to be the probabilistic versions of Guttman (see, e.g., [12]). A relevant example of this stream of literature

is provided by Andrich [13], which supported the statement of Rasch being the natural probabilistic model of the Guttman scale by showing the connection between features from both, such as invariance across persons and items. However, the Rasch model's interpretation as a probabilistic version of the Guttman scale has also been debated on the basis that it is difficult to reconcile the probabilistic nature of the Rasch model with the deterministic nature of a Guttman scale. One of the earliest examples of such critiques can be found in, e.g., Brink [14]. Specifically, the LI assumption that traditionally accompanies Rasch modeling appears to clash with the dependence between the items that are actually at the basis of a Guttman scalogram. As a consequence, it is argued that data patterns that do not conform with a Guttman scale properly fit under Rasch plus LI because the total order established between the difficulties of the items allows modeling such unlikely patterns only because LI is assumed. This suggests that interpreting the Rasch model as a probabilistic version of the Guttman scale might be debatable, since enforcing an order relation among the items by constraining the difficulty parameters is not necessarily the same as setting an order relation between the items themselves. Interestingly, a similar debate can be traced in polytomous IRT models under the name of 'disordered thresholds' or 'reversed deltas' controversy (see, e.g., [15–17]), and originates from the implications of obtaining estimates of the category parameters in a polytomous model which do not follow the order followed by the categories. As, conceptually, a Guttman scalogram is not different to a polytomous IRT model, it is not surprising that the same debate surfaces in both contexts. As it was suggested by Noventa et al. [18], under a KST perspective, there is indeed no formal difference between modeling a set of dependent dichotomous items or a polytomous IRT model. Once the KST concept of knowledge structure is introduced to model the deterministic relation between the basic entities, whether they are items, sub-items, or thresholds capturing the sequence of categories, then probabilistic models can be developed by imposing probability distributions over the states of such structures. Moreover, Noventa et al. [18] suggested that by extending these latter probability distributions to encompass a latent trait, KST models can also be extended to encompass IRT models and provide a generalization of both. We refer here to such an approach as KST-IRT.

The integrated KST-IRT approach is therefore capable of manifesting the definition of Guttman scales using the knowledge structure known as a graded chain (i.e., a chain of items in which the states increase by one item at a time, hence the concept of gradedness). A probabilistic model for the Guttman scale is then obtained by imposing a probability distribution over its states. The probabilities of the response patterns are given by latent class models in which the states of the structure play the role of the latent classes, and their probabilities play the role of the membership probabilities. The conditional error parameters of the model allow for those response patterns which are inconsistent with the assumed knowledge structure. This provides a different approach than the one given by the combination of Rasch plus LI since the latter is not anymore assumed. Most notably, the deterministic response behavior that is induced by the set order relation among items is then decoupled from the total order between the difficulties of the item responses. This also contrasts with the Rasch plus LI approach in which the Rasch model provides a total order between the difficulties of the items.

In Section 2, the IRT, KST, and KST-IRT frameworks are briefly introduced and discussed. The details of model specifications that pertain particularly to KST-IRT models, such as the conditional probability of responses given the knowledge state and the state response functions, are given. In Section 3, frequentist and Bayesian computational approaches are introduced to estimate the item parameters of the KST-IRT models. The setup of the marginalized likelihood for the KST-IRT model and the steps to implement MML-EM and EM-with-Gibbs sampling algorithms are provided. To investigate the performance of the estimation methods and the interplay of IRT and KST-IRT frameworks under local dependence, two simulation studies were conducted and are laid out in Section 4. The first study focuses on the estimation performances of the MML-EM and the EM-with-Gibbs procedures. While the KST-IRT framework allows for specifying the knowledge structures,

two structures that each represent LI and LD were used along with varying the state response functions. For these specifications, the data-generating process and the fitted model were fully interchanged to exhibit model misspecifications. In the second study, the response behavior under local dependence is investigated by comparing the model fit of the Rasch model and KST-IRT models when the item parameters were arranged to generate the graded responses as in the Guttman scale. Deriving from the results, the KST-IRT and IRT models do not provide a symmetric interpretation of the response patterns, especially when the two models are interchanged. Section 5 further details the interpretation of the results in regard to the LD.

2. IRT, KST, and KST-IRT Models

This section provides a brief summary of some IRT and KST notions used in the present manuscript. For more detailed introductions, the readers are referred to comprehensive textbooks, e.g., van der Linden [19] for IRT and Falmagne and Doignon [9] for KST.

2.1. Item Response Models

A general IRT model for the response of the i -th individual to the j -th dichotomous items, with $i \in \{1, \dots, N\}$ and $j \in \{1, \dots, J\}$, is often given in the form of the four parameters logistic model (4PL, [20]), that is

$$P(X_{ij} = 1 | \theta_i, \Gamma_j^4) = c_j + (1 - d_j - c_j) \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}} \quad (1)$$

where $\theta_i \in \mathbb{R}$ expresses the location of the i -th individual along the latent trait and $\Gamma_j^4 = \{a_j, b_j, c_j, d_j\}$ collects the difficulty parameter $b_j \in \mathbb{R}$, the discrimination parameter $a_j \in \mathbb{R}^+$, the guessing parameter $c_j \in [0, 1]$, and the ceiling parameter $d_j \in [0, 1]$ associated to the j -th item. By using a Γ_j^n notation, in which n is the number of item parameters, the Rasch model (RM, [2], also referred to as the 1PL model) is given here by $P(X_{ij} = 1 | \theta_i, \Gamma_j^1)$, the Birnbaum model (also known as the 2PL) is given by $P(X_{ij} = 1 | \theta_i, \Gamma_j^2)$, and the three parameters logistic model (3PL, [21]) is given by $P(X_{ij} = 1 | \theta_i, \Gamma_j^3)$. Typically, the likelihood of the pattern of responses X_i provided by the i -th individual is expressed by means of an assumption of LI, which captures the idea that, conditional on the value of the latent variable, there is no association between the items. A common parametric definition is the (strong) form

$$P(X_i = x_i | \theta_i, \Gamma^n) = \prod_{j=1}^J P(X_{ij} = x_{ij} | \theta_i, \Gamma_j^n) \quad (2)$$

where $\Gamma^n = \{\Gamma_j^n\}$ and the random vector $X_i = [X_{i1}, \dots, X_{iJ}]$ has realizations $x_i = [x_{i1}, \dots, x_{iJ}]$ with $x_{ij} \in \{0, 1\}$. A typical IRT choice to model the probabilistic version of a Guttman scale is thus LI (2) in conjunction with a RM $P(X_{ij} = 1 | \theta_i, \Gamma_j^1)$. In such a case, indeed, the existence of a total order between the item difficulties b_j ensures, by means of LI (2), that those patterns X_i that do not conform to a Guttman scalogram have lower values of their likelihood and are therefore less likely to occur.

2.2. Knowledge State Theory

By means of a combinatorial and set-theoretical approach, KST classifies individuals based on the collections of items that they can master in a given domain of knowledge (see, e.g., [9]). More in detail, let Q be a nonempty set of items q_j (i.e., the domain of knowledge), then a *knowledge state* $K \subseteq Q$ is a collection of problems $q_j \in Q$ that an individual is capable of mastering. A *knowledge structure* is a pair (Q, \mathcal{K}) where \mathcal{K} is a family of subsets of Q that always includes the full domain Q and the empty set \emptyset . As an

example, for the domain $Q = \{q_1, q_2, q_3\}$, a possible knowledge structure is defined by the collection

$$\mathcal{K} = \{\emptyset, \{q_1\}, \{q_2\}, \{q_1, q_2\}, \{q_1, q_3\}, Q\} \quad (3)$$

and represents a situation in which items q_1 and q_2 can be mastered independently of each other, while item q_3 requires item q_1 to be mastered before it can be mastered. Although arbitrary structures $\mathcal{K} \subseteq 2^Q$ can be considered, in what follows, we will only focus on two types of structures, the power set, $\mathcal{K} = 2^Q$, and the graded chain \mathcal{C} . The latter is given by a chain of states that increase one item at a time. As an example, for $Q = \{q_1, q_2, q_3\}$ one has

$$\mathcal{C} = \{\emptyset, \{q_1\}, \{q_1, q_2\}, Q\}. \quad (4)$$

The reason for such a choice is that, as discussed by Noventa et al. [18], in the context of KST-IRT models, the power set represents local independence (see below) while the graded chain represents local dependence. It is indeed evident that the graded chain closely resembles the structural behavior of the Guttman scale (or of a polytomous item) as the strictest form of local dependence where every item is dependent on the previous ones. In order to discuss the KST-IRT approach, first, we need to introduce probabilities over the deterministic structures just defined and discuss two important KST models.

2.2.1. Basic Local Independence Model (BLIM)

A probabilistic knowledge structure (PKS) is a triple (Q, \mathcal{K}, π) in which the deterministic knowledge structure (Q, \mathcal{K}) is made probabilistic by imposing a probability distribution $\pi : \mathcal{K} \rightarrow [0, 1]$ over the knowledge states $K \in \mathcal{K}$. In what follows, we deviate from the traditional KST notation in favor of a more IRT-like notation that allows merging the two better. In order to do so, the symbol X_i is both interpreted as a random vector (as in the IRT case) and as a subset $X_i \in 2^Q$ (as in the KST case). Such bijection between set-theoretic and vector notations was, for instance, discussed by Heller et al. [22] and Noventa et al. [18]. Hence, with a slight abuse of notation, an expression such as $q_j \in X_i$ indicates the fact that the random variable X_{ij} associated with the item q_j takes realization $x_{ij} = 1$ within the random vector X_i , or equivalently that the item q_j belongs to the set $X_i \subseteq Q$. Given this notation, the probability of the response pattern $X_i \in 2^Q$ for the i -th individual is then given by the fundamental equation

$$P(X_i | \Gamma_{\mathcal{K}}) = \sum_{K \in \mathcal{K}} P(X_i | K) \pi(K) \quad (5)$$

where one only needs to set a specific shape for the conditional probabilities $P(X_i | K)$ in order to specify all parameters collected in $\Gamma_{\mathcal{K}}$ and that must be estimated from the data. The most widely used KST model is the BLIM, in which the conditional probabilities $P(X_i | K)$ are written as the product of two sets of parameters

$$P(X_i | K) = \prod_{q \in X_i} \phi_{q,K} \prod_{q' \in Q \setminus X_i} (1 - \phi_{q',K}), \quad (6)$$

where

$$\phi_{q,K} = \begin{cases} 1 - \beta_q & \text{if } q \in K \\ \eta_q & \text{if } q \notin K. \end{cases} \quad (7)$$

where lucky guesses $\eta = \{\eta_q\}$ model correct responses even though an item has not been mastered, whereas careless errors $\beta = \{\beta_q\}$ model incorrect responses even though an item has been mastered. Since $\Gamma_{\mathcal{K}} = \{\pi, \beta, \eta\}$, the BLIM has therefore a total of $|\mathcal{K}| - 1 + 2|Q|$ parameters to be estimated from the data.

2.2.2. Simple Learning Model (SLM)

As mentioned above, the state probabilities π in a PKS are typically treated as a parameter to be estimated with the consequence that large domains Q might imply large structures and, therefore, many state probability parameters. One way to reduce their number is to further constrain the probabilities π . An example is provided by the SLM (see, e.g., [9] (p. 199)), which is used to factorize the probability distributions in learning spaces, which are particularly regular structures that allow for multiple solution strategies while requiring that items are mastered one at a time. The SLM is formally given by

$$\pi(K) = \prod_{q \in K} g_q \prod_{q' \in K^O} (1 - g_{q'}) \quad (8)$$

where $g_q \in]0, 1[$ is a parameter expressing the probability of mastering the item q , and where K^O is called the *outer fringe* of the knowledge state K and is defined as the set of items that can be learned next (one at a time) when being in the knowledge state K , that is

$$K^O := \{q \in Q \setminus K : K \cup \{q\} \in \mathcal{K}\}. \quad (9)$$

In essence, the SLM factorizes the probability $\pi(K)$ of each state (K) into the product of the probabilities of the items that have already been learned (i.e., those in K) and the complementary probabilities of the items that can be learned next (i.e., items in K^O). Generalized versions of the SLM have been explored by Noventa et al. [23].

2.2.3. KST-IRT Models

As formally shown by Noventa et al. [18], if both the state probabilities $\pi(K)$ and the item probabilities g_q in Equations (5) and (8) are extended to encompass some latent variable $\theta \in \mathbb{R}$, so that θ_i expresses the ability of the i -th individual, then the machinery of KST can be used to extend IRT models and to generalize the fundamental Equation (5) and the SLM (8) as

$$P(X_i | \theta_i, \Gamma_{\mathcal{K}}) = \sum_{K \in \mathcal{K}} P(X_i | K) \pi(K | \theta_i) \quad (10)$$

and

$$\pi(K | \theta_i) = \prod_{q \in K} g_q(\theta_i) \prod_{q' \in K^O} (1 - g_{q'}(\theta_i)), \quad (11)$$

respectively, where for sake of simplicity within the terms $\pi(K | \theta_i)$ and $g_q(\theta_i)$ the parameters vector Γ are omitted. If the $g_q(\theta_i)$ functions are identified with the 2PL $P(X_{ij} = 1 | \theta_i, \Gamma_j^2)$ or the RM $P(X_{ij} = 1 | \theta_i, \Gamma_j^1)$, then the state probabilities $\pi(K | \theta_i)$ can be considered *state response functions* that generalize to the knowledge states $K \in \mathcal{K}$, the IRT idea of an item response function. Most of all, two important consequences follow, which are specific to the power set structure $\mathcal{K} = 2^Q$. Firstly, the extended version of the SLM given by Equation (11) provides a generalized version of local stochastic independence in IRT as given by Equation (2) such that, only when the power set 2^Q is considered, then LI (2) is returned. Intuitively, (11) generalizes local independence in that it allows for the presence of items that cannot be mastered from a given state because their prerequisites are not satisfied. Secondly, only when the power set 2^Q is considered, the combination of the extended versions of the SLM and of the BLIM (given by (10) with conditional probabilities set by (6) as in the BLIM) yields exactly the likelihood of a 4PL IRT model as in Equation (1) in the presence of LI (2) in which the following identifications are performed: $c_j = \eta_{q_j}$, $d_j = \beta_{q_j}$, and $g_{q_j}(\theta_i) = P(X_{ij} = 1 | \theta_i, \Gamma_j^2)$ so that $\Gamma_{\mathcal{K}} = \Gamma^4$. These results show that LI (2) is obtained in the KST-IRT approach only in power set case, while if an arbitrary structure is considered, LI does not hold anymore, and one needs to consider the fundamental

Equation (10) (or the generalized version of LI as provided by Equation (11)) to account for the LD captured by the structure. As a consequence, alternatively to the Rasch plus LI approach that is used in IRT, one can capture a Guttman scalogram by considering a graded chain \mathcal{C} as given by Equation (4), together with the latent trait extended fundamental KST equation (10) in conjunction (or not) with the latent trait extended SLM given by Equation (11). Specifically, in what follows, we will compare the IRT approach with two KST-IRT models as provided by Equation (10). The first follows the SLM (11), while the latter assumes a different shape of the state response functions $\pi(K)|\theta_i$ and is discussed in the following subsection.

2.2.4. Logistic Knowledge Structure

A perfect example of a KST-IRT model is provided by a Logistic Knowledge Structure ([24], LKS). If one indeed considers an arbitrary knowledge structure \mathcal{K} , one can define an LKS as a 5-tuple $(Q, \mathcal{K}, b, P, \pi)$ where $b : Q \rightarrow \mathbb{R}$, $\pi : \mathcal{K} \times \mathbb{R} \rightarrow (0, 1)$ are such that

$$P(X_i|\theta_i) = \sum_{K \in \mathcal{K}} P(X_i|K)\pi(K|\theta_i) = \sum_{K \in \mathcal{K}} P(X_i|K) \frac{\exp \left[\sum_{q_j \in K} (\theta_i - b_{q_j}) \right]}{\sum_{L \in \mathcal{K}} \exp \left[\sum_{q_j \in L} (\theta_i - b_{q_j}) \right]} \quad (12)$$

where $P(X_i|K)$ is given by the BLIM (6). The LKS can be seen as a generalization of the partial credit model (PCM) to an arbitrary structure rather than a graded chain [18]. Finally, notice that while the SLM approach of Equation (11) resembles an IRT sequential/step model, the LKS of Equation (12) resembles a polytomous IRT model. For this reason, they provide two different ways of modeling LD that are considered in what follows.

3. Estimation Methods

Since the introduction of the MML [25] and MML-EM [26] that remedy the inconsistent parameter estimates of the joint maximum likelihood estimation (JMML, [21]), the MML-EM has served as a standard method to estimate the parameters of many psychometric models [27]. In order to separate the item parameters from the ability parameter, the joint distribution is marginalized over a known examinee ability distribution.

One known shortcoming of the MML-EM is that when this distributional assumption is violated, the estimates can produce undesirably deviating results [28,29]. To anticipate the deviation, one can restrict the estimates within a reasonable range by imposing a known distribution for the item parameters. This Bayesian approach [30,31] was further developed by incorporating the Markov chain Monte Carlo (MCMC) simulation with the Bayesian estimation (Metropolis-Hastings, Robbins-Monro, [32,33] blocked Metropolis algorithm [34,35]). MML-EM and the MCMC methods are generally more computationally intensive than the conditional maximum likelihood estimation [36] used for the RM. However, they allow for more flexible and accurate estimation of the parameters in more complex models such as KST-IRT models.

To this end, two estimation methods, the MML-EM, and a Bayesian method, EM-with-Gibbs, were used to recover the item parameters of the KST-IRT models. The item parameters to be estimated are, per each item j , the BLIM parameters, η_j and β_j , and the item difficulties b_j . In the first subsection, we derive the marginalized likelihood for the KST-IRT models and define the necessary parts uniquely facilitated in the MML estimation. In the following subsections, the steps for implementing each MML-EM and the EM-with-Gibbs sampling method were presented.

3.1. Marginalized Likelihood for KST-IRT Models

Following the same notation used in Section 2, we denote by Γ the collection of all item parameters and by $\zeta \in \Gamma$ an arbitrary target parameter. To recover the item parameters from the response patterns utilizing the marginalized likelihood, the conducted estimation

procedure resembles that of Bock and Aitkin [26]. Under the KST-IRT model, the conditional probability of the response pattern X_i is given by (10). For an examinee sampled from a population with a known distribution, $f(\theta)$, the unconditional probability of the response pattern, X_i , for the i -th individual is given by:

$$P(X_i|\Gamma) = \int P(X_i|\theta_i, \Gamma) f(\theta_i) d\theta_i. \quad (13)$$

The log-likelihood is denoted $\log L(X|\Gamma) = \sum_i \log P(X_i|\Gamma)$ where $X = [X_1, \dots, X_N]$ is the full data matrix. Then, to derive the marginal likelihood, we take the first derivative of the log-likelihood with regard to the target parameter ξ . By using the Bayes' rule that $f(\theta_i|X_i, \Gamma) = \frac{P(X_i|\theta_i, \Gamma) f(\theta_i)}{P(X_i|\Gamma)}$ and using the relation

$$\frac{\partial}{\partial \xi} P(X_i|\theta_i, \Gamma) = \frac{\partial}{\partial \xi} [\log P(X_i|\theta_i, \Gamma)] P(X_i|\theta_i, \Gamma), \quad (14)$$

we have the following convenient form where the posterior distribution, $f(\theta_i|X_i, \Gamma)$, replaces the prior distribution of the ability parameter, $f(\theta_i)$, in the formula:

$$\begin{aligned} \frac{\partial}{\partial \xi_j} \log L(X|\Gamma) &= \sum_i^N \frac{1}{P(X_i|\Gamma)} \int \left[\frac{\partial}{\partial \xi_j} P(X_i|\theta_i, \Gamma) \right] f(\theta_i) d\theta_i \\ &= \sum_i^N \int \left[\frac{\partial}{\partial \xi_j} \log P(X_i|\theta_i, \Gamma) \right] f(\theta_i|X_i, \Gamma) d\theta \end{aligned} \quad (15)$$

The next step is to attain the first derivative of $P(X_i|\theta_i, \Gamma)$. Let x_{ij} denote the dichotomous response of individual i to item j , and let $P(x_{ij}|\theta_i, \Gamma_j^n)$ and $Q(x_{ij}|\theta_i, \Gamma_j^n)$ denote the probabilities of correct and incorrect responses given θ_i , then by LI (2) we can write

$$\begin{aligned} \frac{\partial}{\partial \xi_j} \log P(X_i|\theta_i, \Gamma) &= \frac{\partial}{\partial \xi_j} \log \left[\prod_j^J P(x_{ij}|\theta_i, \Gamma_j^n)^{x_{ij}} Q(x_{ij}|\theta_i, \Gamma_j^n)^{1-x_{ij}} \right] \\ &= \sum_j^J \frac{x_{ij} - P(x_{ij}|\theta_i, \Gamma_j^n)}{P(x_{ij}|\theta_i, \Gamma_j^n) Q(x_{ij}|\theta_i, \Gamma_j^n)} \frac{\partial P(x_{ij}|\theta_i, \Gamma_j^n)}{\partial \xi_j}. \end{aligned} \quad (16)$$

So far, the above procedure provided by Bock and Lieberman [25] for decomposing the log-likelihood is common in IRT. However, with KST-IRT models, $P(X_i|\theta_i, \Gamma_K)$ as given by the fundamental equation (10) expands accordingly to the specified knowledge structure. Therefore, Equation (16) needs to be replaced by the corresponding element of the Jacobian matrix, which now depends on both the choice of a structure \mathcal{K} and state response probability $\pi(K|\theta_i)$. Indeed, in attaining the first derivatives of the difficulty item parameters, one would obtain different results for the SLM and the LKS. Similarly, the Jacobian matrix has different sizes for different structures, i.e., $(J+1) \times (3J)$ for a graded chain and $(2^J) \times (3J)$ for a power set. To illustrate these differences, let us consider a set of two items, $Q = \{q_1, q_2\}$ where graded chain yields $\mathcal{C} = \{\emptyset, \{q_1\}, Q\}$ and the power set yields $2^Q = \{\emptyset, \{q_1\}, \{q_2\}, Q\}$. Let then x_{ij} denote, as usual, the binary response for a choice of the item response functions $g_{q_j}(\theta_i) = P(X_{ij} = 1|\theta_i, \Gamma_j^1)$ as in the RM, and let $h_{q_j}(\theta_i) = P(X_{ij} = 0|\theta_i, \Gamma_j^1)$ denote the response function for the incorrect response. Then the first derivatives of each item parameter for item $j = 1$, when the state response function is modeled according to the SLM (11), yields for the graded chain \mathcal{C}

$$\frac{\partial}{\partial \xi_1} P_{\text{SLM}}(X_i|\theta_i, \Gamma_C) = \begin{cases} (-1)^{x_{i1}+1} \eta_{q_2}^{x_{i2}} (1 - \eta_{q_2})^{1-x_{i2}} h_{q_1}(\theta_i) & \xi = \eta \\ (-1)^{x_{i1}} (1 - \beta_{q_2})^{x_{i2}} \beta_{q_2}^{1-x_{i2}} g_{q_1}(\theta_i) g_{q_2}(\theta_i) & \xi = \beta \\ \left[P(X_i|\emptyset) - P(X_i|\{q_1\}) h_{q_2}(\theta_i) \right. \\ \left. - P(X_i|Q) g_{q_2}(\theta_i) \right] g_{q_1}(\theta_i) h_{q_1}(\theta_i) & \xi = b \end{cases}$$

and for the power set 2^Q

$$\frac{\partial}{\partial \xi_1} P_{\text{SLM}}(X_i|\theta_i, \Gamma_{2^Q}) = \begin{cases} (-1)^{x_{i1}+1} h_{q_1}(\theta_i) \left[\eta_{q_2}^{x_{i2}} (1 - \eta_{q_2})^{1-x_{i2}} h_{q_2}(\theta_i) \right. \\ \left. + (1 - \beta_{q_2})^{x_{i2}} \beta_{q_2}^{1-x_{i2}} \right] & \xi = \eta \\ (-1)^{x_{i1}} g_{q_1}(\theta_i) \left[\eta_{q_2}^{x_{i2}} (1 - \eta_{q_2})^{1-x_{i2}} h_{q_2}(\theta_i) \right. \\ \left. + (1 - \beta_{q_2})^{x_{i2}} \beta_{q_2}^{1-x_{i2}} g_{q_2}(\theta_i) \right] & \xi = \beta \\ g_{q_1}(\theta_i) h_{q_1}(\theta_i) \left[(P(X_i|\emptyset) h_{q_2}(\theta_i) - P(X_i|\{q_1\}) g_{q_2}(\theta_i)) \right. \\ \left. - (P(X_i|\{q_1\}) h_{q_2}(\theta_i) - P(X_i|Q) g_{q_2}(\theta_i)) \right] & \xi = b \end{cases}$$

where $P_{\text{SLM}}(X_i|\theta_i, \Gamma_C)$ and $P_{\text{SLM}}(X_i|\theta_i, \Gamma_{2^Q})$ take the form of Equation (10). Specifically, the first term of each sum in Equation (10), $P(X_i|\mathcal{K})$ contains η and β in the form of BLIM (6) and the second term corresponding to (11) contains the difficulty parameters, b , where the factorized g_q follows the form of RM.

Similarly, the first derivatives of each item parameter for item $j = 1$, when the state response function is modeled according to the LKS (12), yields for the graded chain \mathcal{C}

$$\frac{\partial}{\partial \xi_1} P_{\text{LKS}}(X_i|\theta_i, \Gamma_C) = \begin{cases} (-1)^{x_{i1}} \eta_{q_2}^{x_{i2}} (1 - \eta_{q_2})^{1-x_{i2}} \pi(\emptyset|\theta_i) & \xi = \eta \\ (-1)^{x_{i1}} (1 - \beta_{q_2})^{x_{i2}} \beta_{q_2}^{x_{i2}} \pi(Q|\theta_i) & \xi = \beta \\ P(X|\emptyset) (\pi(\emptyset|\theta_i) (\pi(q_1|\theta_i) + \pi(Q|\theta_i))) \\ - P(X_i|q_1) (\pi(\emptyset|\theta_i) \pi(q_1|\theta_i)) & \xi = b \\ - P(X_i|Q) (\pi(\emptyset|\theta_i) \pi(Q|\theta_i)) \end{cases}$$

and for the power set 2^Q

$$\frac{\partial}{\partial \xi_1} P_{\text{LKS}}(X_i|\theta_i, \Gamma_{2^Q}) = \begin{cases} (-1)^{x_{i1}+1} \left[\eta_2^{x_{i2}} (1 - \eta_2)^{1-x_{i2}} \pi(\emptyset|\theta_i) + (1 - \beta_2)^{x_{i2}} \beta_2^{1-x_{i2}} \pi(q_2|\theta_i) \right] & \xi = \eta \\ (-1)^{x_{i1}} \left[\eta_2^{x_{i2}} (1 - \eta_2)^{1-x_{i2}} \pi(\emptyset|\theta_i) + (1 - \beta_2)^{x_{i2}} \beta_2^{1-x_{i2}} \pi(Q|\theta_i) \right] & \xi = \beta \\ \left[(P(X_i|\emptyset) \pi(\emptyset|\theta_i) + P(X_i|q_2) \pi(q_2|\theta_i)) (\pi(q_1|\theta_i) \pi(Q|\theta_i)) \right. \\ \left. - (P(X_i|q_1) \pi(q_1|\theta_i) + P(X_i|Q) \pi(Q|\theta_i)) (\pi(\emptyset|\theta_i) + \pi(q_2|\theta_i)) \right] & \xi = b \end{cases}$$

where the second term of each sum in $P_{\text{LKS}}(X_i|\theta_i, \Gamma_C)$ and $P_{\text{LKS}}(X_i|\theta_i, \Gamma_{2^Q})$ now correspond to the divide-by-total approach (12).

As a final step, for marginalizing the likelihood over the ability parameter, the integral in (16) is replaced with a summation. Bock and Lieberman [25] used numerical quadrature as an approximation for the integration. The assumed continuous density for the ability parameter is filled with rectangles of equal widths to be summed over, and the midpoints of the rectangles are nodes denoted as $Y_k, k \in \{1, \dots, s\}$, and each node is associated with a quadrature weight, $A(Y_k)$. Typically, a normal density is applied for the distribution of the individual ability parameter. We can now compute the posterior distribution of the finite partitions of θ at each node,

$$P(Y_k|X) = \frac{\sum_K P(X|K) \pi(K|\theta) A(Y_k)}{\sum_k \sum_K P(X|K) \pi(K|\theta) A(Y_k)}. \tag{17}$$

Similarly to Bock and Lieberman [25], x_{ij} and $P(x_{ij}|\theta_i, \Gamma_j^n)$ in Equation (16) are replaced by

$$\bar{n}_k = \sum_i^N P(Y_k|X_i) = \sum_i^N \left[\frac{\sum_K P(X_i|K)\pi(K|\theta_i)A(Y_k)}{\sum_K^q \sum_K P(X_i|K)\pi(K|\theta_i)A(Y_k)} \right] \quad (18)$$

and

$$\bar{r}_{jk} = \sum_i^N x_{ij}P(Y_k|X_i) = \sum_i^N \left[\frac{x_{ij} \sum_K P(X_i|K)\pi(K|\theta)A(Y_k)}{\sum_K^q \sum_K P(X_i|K)\pi(K|\theta)A(Y_k)} \right], \quad (19)$$

respectively. These two new artificial measurements given by the observed responses act as sufficient statistics as they represent the expected number of individuals at ability level Y_k and the expected number of correct responses at ability level Y_k for item j , respectively.

3.2. MML-EM

Estimating the item parameters with the MML-EM involves iterative computation of expectation and maximization. The following two-step procedure is repeated until a predetermined convergence criterion is met:

E-step: with the response patterns X , let $\eta^{(0)}$, $\beta^{(0)}$, and $b^{(0)}$ be the collections of initial parameters, we compute on each node Y_k the expected number \bar{n}_k of individuals and the expected number of correct responses \bar{r}_{jk} for item j .

M-step: we solve the first derivatives for each item via a numerical method such as Newton-Raphson's.

In the expectation step, the initial parameters for the BLIM, $\eta^{(0)}$ and $\beta^{(0)}$, can be drawn from a uniform distribution with a range of]0.05, 0.4[. The initial values for the difficulty parameters $b_j^{(0)}$ can be drawn from a uniform distribution with a range of] - 3, 3[. For the prior distribution of the ability parameter, we consider a standard normal distribution.

In the maximization step, the parameters are updated to maximize the expected log likelihood. Once the difficulty parameters are estimated, the solutions to the first derivatives can be attained for each item parameter independently as the artificial quantities \bar{n}_k and \bar{r}_{jk} provide sufficient statistics for the parameter estimates. As each iteration of the MML-EM can be solved quickly with the closed-form first derivatives of the marginalized likelihoods, the short estimation time is the advantage of this method.

3.3. EM-with-Gibbs

The Gibbs sampling technique is an MCMC algorithm that draws samples from full conditional distributions. As a special case of Metropolis-Hastings, it leverages known prior distributions to approximate the posterior distribution. This approach can be viewed as a generalized EM algorithm, where the standard expectation step is replaced by drawing samples from conditional distributions. In particular, when the distribution of item parameters is known, the Gibbs sampling procedure can be effectively combined with the EM algorithm.

The present study proposes an approach for estimating the parameters of the KST-IRT models by combining the Gibbs sampling method with the EM algorithm. This approach differs from the MML-EM in that the BLIM parameters are drawn from the full conditional distributions. The log-likelihoods of the response patterns are computed with the drawn sample instead of using the more rigidly computed artificial measurements as (18) and (19) as in the MML-EM. The resulting parameter samples are subsequently used to update the joint distribution in the next iteration of the algorithm.

The difficulty parameter is constrained to reflect the response behavior of the graded chain in the IRT case, meaning that the difficulty parameters are strictly increasing. To enforce this constraint, a multivariate normal distribution is used with means that impose the order of increasing difficulty parameters. For the BLIM parameters, which are supported on the interval]0, 1[, a Beta (1,1) flat prior can be assumed. Once the frequencies of guessing

and careless error incidences are incorporated into the prior distribution, the resulting posterior distribution with updated parameters can be used to draw estimates of η and β . In the likelihoods that follow, \hat{K}_i^{\max} represents the knowledge state that maximizes $\pi(K|\theta_i)$ for the i -th individual given the response pattern X :

$$\begin{aligned} L(\eta|X_i, \beta, b) &\propto \eta_{q_j}^{\sum_i^N x_{ij}[q_j \notin \hat{K}_i^{\max}]} (1 - \eta_{q_j})^{\sum_i^N (1-x_{ij})[q_j \notin \hat{K}_i^{\max}]} \\ L(\beta|X_i, \eta, b) &\propto (1 - \beta_{q_j})^{\sum_i^N x_{ij}[q_j \in \hat{K}_i^{\max}]} \beta_{q_j}^{\sum_i^N (1-x_{ij})[q_j \in \hat{K}_i^{\max}]} \end{aligned} \quad (20)$$

where $[q_j \notin \hat{K}_i^{\max}]$ is the Iverson brackets that attain value one when the logical condition is satisfied and zero otherwise. Normalizing the conditional likelihoods in (20), we attain the Beta distribution where the BLIM parameters can be drawn from, that is

$$\begin{aligned} \eta_{q_j} &\sim \text{Beta}\left(\sum_i^N x_{ij}[q_j \notin \hat{K}_i^{\max}], \sum_i^N [q_j \notin \hat{K}_i^{\max}]\right) \\ \beta_{q_j} &\sim \text{Beta}\left(\sum_i^N (1-x_{ij})[q_j \in \hat{K}_i^{\max}], \sum_i^N [q_j \in \hat{K}_i^{\max}]\right). \end{aligned} \quad (21)$$

Due to the sampling nature of the Gibbs method, the iterations are not stopped by convergence criteria. Instead, a fixed number of iterations are repeated as follows:

Step 1: with the response patterns, X , values θ_i are sampled from the posterior distribution of the ability parameter given by the computed nodes $A(Y_k)$ for some fine grid $k \in \{1, \dots, s\}$ with the vectors of initial item parameters $\eta^{(0)}$, $\beta^{(0)}$, and $b^{(0)}$.

Step 2: draw a predetermined batch size of difficulties b from a multivariate normal distribution with a mean vector $\mu = [\mu_1, \dots, \mu_j]$ that previously maximized the joint likelihood.

Step 3: draw from the conditional likelihood of the BLIM parameters and update the current estimate.

In the first step, the use of the grid of θ and sampling of the node that maximizes the likelihood of the response patterns act as the computation of the expected number of individuals at each node in the MML-EM algorithm. For the difficulty parameter, the sampling procedure differs from those of the BLIM parameters that are drawn from a full conditional distribution as suggested by the traditional Gibbs sampling. Specifically, difficulty parameters are not drawn from each one from their own distribution, but they are instead drawn from a multivariate distribution so that they are used as an entirety of J -tuple vector for computing the state response function. Given an informative prior, it is more sensible to draw a batch of J -tuple vectors and select the one that returns the maximum likelihood at each draw. Then, the BLIM parameters are drawn with the standard Gibbs sampling method, where the Beta distribution parameters are updated at each iteration. Finally, in MCMC simulations, the burn-in period is the number of initial iterations that are discarded to ensure that the chain has converged to the stationary distribution. During the burn-in period, the chain may be in a transient state, which means that it has not yet reached its equilibrium distribution. In this application, 1000 iterations were used to determine the burn-in period.

4. Simulation Studies

In the first simulation study, we evaluate the efficacy of the parameter estimation methods by comparing the recovery performance of KST-IRT model parameters. Specifically, we compare the estimates and the root mean squared error (RMSE) of the KST-IRT model parameters using different knowledge structures and state response functions. A further comparison of the estimation performance is demonstrated by misspecifying the assumed knowledge structure (i.e., either a graded chain or a power set) and/or the state response function (i.e., either a SLM or a LKS). In the second simulation study, we examine

the model fit of the RM and the KST-IRT models when the data-generating process and fitted model are interchanged. In other words, the response pattern generated from RM is fitted with KST-IRT models and vice versa. As RM and KST-IRT models, as discussed in Section 2, assume completely different processes underlying the generation of unlikely data patterns (i.e., non-Guttman patterns of responses), this latter simulation study allows for assessing the models' ability to capture the underlying structure of the data and to understand which form of modeling of a Guttman scale is more robust to violations of its own assumptions.

The true item parameters were established to mimic a Guttman scalogram. In both simulation studies, the difficulty parameters of the items are arranged in increasing order. This arrangement was chosen to generate responses that are consistent with the assumed characteristic of the Guttman scale according to the IRT perspective, where each successive item is expected to be more difficult than the previous one. In comparison, the KST-IRT model explicitly accounts for local dependence of the items through the knowledge structure (i.e., the graded chain), and for such a reason, it would not need to a priori establish an increasing order for the difficulties. Nonetheless, the assumption was taken to meet the usual requirements of IRT specifications. It should also be stressed that, since the condition of LI is obtained in KST-IRT models when the knowledge structure is a power set, the comparison between RM plus LI vs SLM plus power set actually corresponds to comparing a Rasch model as it is traditionally applied in IRT with a 4PL model in which the Birnbaum parameter is set to one.

In the two simulation studies, a total of five models were considered: one IRT model, the Rasch Model (RM), and four KST-IRT models. The KST-IRT models are identified by their respective knowledge structures and state response functions. Each KST-IRT model is labeled by the assumed knowledge structure: either a graded chain denoted as \mathcal{C} or a power set denoted as 2^Q where Q represents the full domain. Additionally, the choice of the state response function is indicated as either SLM or LKS. In order to simplify the interpretations of the models for the reader, we summarize here which different combinations actually correspond.

- *RM*: this is the traditional assumption of IRT models in which a Rasch model is considered in conjunction with the assumption of LI.
- $\mathcal{C} + \text{SLM}$: this is the first KST-IRT version of a Guttman scale in which a graded chain is used to model the scalogram, and the state response functions are factorized by means of the generalized version of LI provided by Equation (11). The model is structurally similar to a sequential/step IRT model. The conditional error parameters provide the noise generating the non-Guttman pattern of responses.
- $\mathcal{C} + \text{LKS}$: this is a second KST-IRT version of a Guttman scale in which a graded chain is used to model the scalogram, but this time the state response functions are not factorized. Since a LKS (12) is used, the state response function rather takes the form of a polytomous IRT model applied to the Guttman scale. As previously, the conditional error parameters provide the noise generating the non-Guttman pattern of responses.
- $2^Q + \text{SLM}$: this is the KST-IRT version of a 4PL model plus LI (but the discrimination parameters are set to one).
- $2^Q + \text{LKS}$: by construction, this model is equivalent to the former, so it was not reported in the results of the second simulation study since the estimates are essentially the same, and the models generate essentially the same data.

In both simulation studies, each design was repeated $N_{rep} = 50$ times to compute the average of the estimates and the RMSE to evaluate the dispersion of the estimate distribution for a target parameter ξ_j . Preliminary experiments with smaller and larger set of replications indicated that augmenting the number of replications beyond 50 did not yield a substantial reduction in their respective RMSE or significantly enhance the precision of the estimates. Consequently, $N_{rep} = 50$ was adopted as a reasonable compromise

between acquiring reliable results and managing computational burden. The RMSE was calculated with the following formula:

$$RMSE = \sqrt{\frac{\sum_k^{N_{rep}} (\xi_j - \hat{\xi}_{jk})^2}{N_{rep}}} \quad (22)$$

where $\hat{\xi}_{jk}$ represents the k -th replication of the parameter estimate.

The statistical computations in this study were executed using the programming language R and its extension, Rcpp, which utilizes C++. To compute the difficulty parameters for the RM estimates, we employed the *rasch* function from the *ltm* (version 1.2-0) package. For the estimation of the KST-IRT parameters, b , η , and β , the two introduced estimation procedures, MML-EM and EM-with-Gibbs, were implemented in R. The simulations and computations were performed on a computing system with an Intel Core i7-10610U CPU, operating at a speed of 1.80 GHz to 2.30 GHz and 32 GB of random access memory (RAM).

The simulation time for both MML-EM and EM-with-Gibbs is heavily influenced by the length of the assessment and the knowledge structure. It is vital to note that computation time can escalate rapidly as the number of items increases. Among the dimensions of experimentation in the simulation, the most significant contrast in computation time was observed between the estimation methods. The computation time of MML-EM is primarily contingent upon the duration of Newton-Raphson iterations, while EM-with-Gibbs is contingent upon the specified run length of the chain. Specifically, for the simulation, the average computation times of MML-EM under SLM and LKS were 0.87 and 0.74 h, respectively. Correspondingly, the average computation times of EM-with-Gibbs under SLM and LKS were 1.24 and 1.23 h.

4.1. Simulation Study 1

The objective of this study is to assess the accuracy of item parameter recovery by means of the MML-EM and EM-with-Gibbs sampling methods under correct and incorrect model specifications. To this end, we generated response patterns for $N = 10,000$ individuals with θ randomly sampled from a standard normal distribution. Each individual responded to $J = 5$ items, with item parameters assumed as follows: for the difficulty parameters, $b = (-2, -1, 0, 1, 2)$, for the lucky guess parameter, $\eta = (0.129, 0.100, 0.074, 0.115, 0.096)$ and careless error parameters, $\beta = (0.158, 0.112, 0.063, 0.148, 0.147)$. The state response function, $\pi(K|\theta_i)$, was computed based on two knowledge structures: power set, 2^Q , and graded chain, \mathcal{C} , where η_{q_j} and β_{q_j} correspond to the guessing and slipping parameters. For the power set 2^Q , Q denotes the full domain where the considered domain for this simulation is $Q = \{q_1, q_2, q_3, q_4, q_5\}$ for $J = 5$.

A $2 \times 2 \times 2$ design was conducted using two computational methods, MML-EM and EM-with-Gibbs, two state response functions, SLM and LKS, and two knowledge structures, \mathcal{C} and 2^Q , where each of the five items has three item parameters, i.e., $\Gamma_j = \{\eta_{q_j}, \beta_{q_j}, b_j\}$.

Parameter recovery under correctly specified knowledge structure and state response probability is presented in Table 1. The first column indicates the target parameter ξ_j for which we provide the mean estimated values, the second column specifies the state response probability $\pi(K|\theta)$, and the third column specifies the knowledge structure \mathcal{K} which is paired with the state response probability. Then, every block of five columns contains the mean estimates and RMSE of each parameter \hat{b}_j , $\hat{\eta}_{q_j}$, and $\hat{\beta}_{q_j}$, for all the items $j \in \{1, \dots, 5\}$. The two blocks refer to the two estimation methods, MML-EM and EM-with-Gibbs, respectively. In each cell, the first row is the arithmetic mean of the estimates for each condition, and the second in parentheses is the RMSE.

Table 1. Parameter estimation of KST-IRT models under correctly specified structure \mathcal{K} and state response probability $\pi(K|\theta)$.

| ξ_j | $\pi(K \theta)$ | \mathcal{K} | MML-EM | | | | | EM-with-Gibbs | | | | |
|---------------|-----------------|---------------|-----------------|-----------------|-----------------|----------------|----------------|-----------------|-----------------|----------------|----------------|----------------|
| | | | j | | | | | j | | | | |
| | | | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| b_j | SLM | \mathcal{C} | -1.94 (0.07) | -1.05 (0.06) | 0.05 (0.06) | 1.03 (0.04) | 1.94 (0.05) | -2.03 (0.06) | -1.04 (0.05) | 0.02 (0.03) | 1.03 (0.05) | 2.04 (0.06) |
| | | 2^Q | -1.97 (0.08) | -1.05 (0.06) | 0.02 (0.04) | 1.02 (0.07) | 1.95 (0.04) | -2.03 (0.03) | -1.02 (0.03) | 0.03 (0.06) | 1.03 (0.04) | 2.03 (0.04) |
| | LKS | \mathcal{C} | -1.95 (0.06) | -0.99 (0.02) | -0.02 (0.04) | 1.04 (0.05) | 1.96 (0.06) | -1.96 (0.07) | -0.98 (0.05) | 0.03 (0.04) | 0.96 (0.05) | 1.98 (0.03) |
| | | 2^Q | -1.95 (0.08) | -1.05 (0.06) | 0.05 (0.05) | 1.04 (0.05) | 1.98 (0.06) | -1.97 (0.07) | -1.05 (0.03) | 0.02 (0.05) | 1.04 (0.04) | 1.96 (0.04) |
| η_{q_j} | SLM | \mathcal{C} | 0.10 (0.06) | 0.13 (0.05) | 0.10 (0.03) | 0.14 (0.04) | 0.12 (0.05) | 0.15 (0.04) | 0.12 (0.04) | 0.06 (0.03) | 0.10 (0.03) | 0.07 (0.02) |
| | | 2^Q | 0.16 (0.07) | 0.13 (0.05) | 0.09 (0.04) | 0.13 (0.05) | 0.11 (0.05) | 0.08 (0.04) | 0.05 (0.04) | 0.14 (0.05) | 0.08 (0.02) | |
| | LKS | \mathcal{C} | 0.11 (0.06) | 0.12 (0.05) | 0.09 (0.05) | 0.10 (0.04) | 0.13 (0.05) | 0.11 (0.04) | 0.12 (0.03) | 0.06 (0.03) | 0.10 (0.04) | 0.08 (0.02) |
| | | 2^Q | 0.11 (0.03) | 0.08 (0.02) | 0.10 (0.03) | 0.13 (0.05) | 0.11 (0.04) | 0.12 (0.05) | 0.09 (0.05) | 0.09 (0.02) | 0.10 (0.04) | 0.11 (0.02) |
| β_{q_j} | SLM | \mathcal{C} | 0.12 (0.06) | 0.13 (0.04) | 0.08 (0.05) | 0.13 (0.03) | 0.12 (0.03) | 0.19 (0.03) | 0.08 (0.01) | 0.09 (0.02) | 0.12 (0.03) | 0.14 (0.04) |
| | | 2^Q | 0.13 (0.03) | 0.09 (0.02) | 0.04 (0.04) | 0.18 (0.04) | 0.17 (0.03) | 0.18 (0.02) | 0.10 (0.02) | 0.05 (0.02) | 0.13 (0.03) | 0.16 (0.03) |
| | LKS | \mathcal{C} | 0.13 (0.02) | 0.09 (0.04) | 0.04 (0.01) | 0.15 (0.02) | 0.16 (0.03) | 0.18 (0.02) | 0.13 (0.02) | 0.05 (0.01) | 0.15 (0.02) | 0.13 (0.02) |
| | | 2^Q | 0.13 (0.03) | 0.09 (0.04) | 0.09 (0.04) | 0.17 (0.05) | 0.17 (0.03) | 0.13 (0.03) | 0.09 (0.01) | 0.05 (0.02) | 0.17 (0.02) | 0.13 (0.02) |

Note. Each cell contains the average of the simulated estimates and the RMSE of the estimates in the parenthesis. b_j , η_{q_j} , and β_{q_j} represent the difficulty, lucky guess, and careless error parameters for item q_j .

In Table 1, the performance of the MML-EM and the EM-with-Gibbs can be compared in terms of bias and the RMSE. A general finding was that the MML-EM yielded higher absolute bias and RMSE than the EM-with-Gibbs for all parameters. Between SLM and LKS, using the LKS for the state probabilities showed an overall smaller bias than those of the SLM. However, the difference is not too apparent. The absolute bias of b , η , and β were, in general, less than 0.06, 0.035, and 0.04, respectively. The bias and the RMSE behaved similarly between the two knowledge structures, \mathcal{C} and 2^Q .

Tables 2 and 3 display instead the parameter estimation under misspecification of both knowledge structure \mathcal{K} and state response function $\pi(K|\theta)$, each for MML-EM and EM-with-Gibbs sampling, respectively. Misspecifying both the knowledge structure and the state response function, this simulation investigated the full misspecification of the two components of the KST-IRT model. The first column indicates the target parameter ξ_j being estimated. The second and third columns specify the state response probability $\pi(K|\theta)$ used for data generation (DG) and fitted model (FM). The subsequent fourth and fifth columns indicate the knowledge structure \mathcal{K} used for data generation and the fitted model. Starting from the first row, every two consecutive rows of two-way mismatched knowledge structures pertain to a paired mismatch of the state response function.

Table 2. Parameter estimation of KST-IRT models under fully misspecified structure \mathcal{K} (data-generating (DG) model and fitted model (FM) and state response probability $\pi(K|\theta)$ with MML-EM estimation.

| ξ_j | $\pi(K \theta)$ | | \mathcal{K} | | MML-EM | | | | |
|---------------|-----------------|-----|---------------|---------------|-----------------|-----------------|----------------|----------------|----------------|
| | | | | | j | | | | |
| | DG | FM | DG | FM | 1 | 2 | 3 | 4 | 5 |
| b_j | SLM | LKS | 2^Q | \mathcal{C} | −1.90 (0.19) | −1.10 (0.13) | 0.04 (0.05) | 1.13 (0.15) | 1.83 (0.21) |
| | | | \mathcal{C} | 2^Q | −2.21 (0.31) | −1.15 (0.18) | 0.03 (0.04) | 0.85 (0.16) | 1.99 (0.12) |
| | LKS | SLM | 2^Q | \mathcal{C} | −1.88 (0.21) | −1.16 (0.13) | 0.07 (0.06) | 1.14 (0.20) | 1.80 (0.22) |
| | | | \mathcal{C} | 2^Q | −2.24 (0.38) | −1.22 (0.21) | 0.06 (0.08) | 0.84 (0.16) | 2.12 (0.19) |
| η_{q_j} | SLM | LKS | 2^Q | \mathcal{C} | 0.14 (0.04) | 0.10 (0.03) | 0.05 (0.02) | 0.13 (0.04) | 0.09 (0.03) |
| | | | \mathcal{C} | 2^Q | 0.15 (0.05) | 0.12 (0.04) | 0.06 (0.04) | 0.06 (0.03) | 0.09 (0.03) |
| | LKS | SLM | 2^Q | \mathcal{C} | 0.15 (0.06) | 0.09 (0.04) | 0.09 (0.03) | 0.09 (0.04) | 0.09 (0.04) |
| | | | \mathcal{C} | 2^Q | 0.13 (0.05) | 0.12 (0.04) | 0.07 (0.02) | 0.10 (0.02) | 0.11 (0.04) |
| β_{q_j} | SLM | LKS | 2^Q | \mathcal{C} | 0.18 (0.05) | 0.14 (0.04) | 0.03 (0.02) | 0.15 (0.03) | 0.14 (0.03) |
| | | | \mathcal{C} | 2^Q | 0.15 (0.04) | 0.15 (0.06) | 0.10 (0.03) | 0.18 (0.05) | 0.13 (0.05) |
| | LKS | SLM | 2^Q | \mathcal{C} | 0.17 (0.03) | 0.08 (0.03) | 0.03 (0.01) | 0.12 (0.04) | 0.15 (0.02) |
| | | | \mathcal{C} | 2^Q | 0.14 (0.03) | 0.10 (0.03) | 0.04 (0.02) | 0.13 (0.04) | 0.16 (0.04) |

Note. Each cell contains the average of the simulated estimates and the RMSE of the estimates in the parenthesis. b_j , η_{q_j} , and β_{q_j} represent the difficulty, lucky guess, and careless error parameters for item q_j .

Comparing Tables 2 and 3, the estimates given by EM-with-Gibbs approximate closer to the true parameter values with relatively lower RMSE than those of MML-EM. It is, however, difficult to conclude that the estimates of the EM-with-Gibbs are superior to those of MML-EM in all cases. In terms of bias for the difficulty parameter for all four designs, 16 cells out of 20 of Table 3 have a smaller bias magnitude than those of Table 2 when we compute the between-item RMSE for the difficulty parameter, for instance, Tables 2 and 3 each yield 0.14 and 0.12. The overall bias and RMSE are greater than the case of correctly specified state response probability $\pi(K|\theta)$ and knowledge structure \mathcal{K} for both estimation methods. The Gibbs sampling approach consistently performs better than the MML-EM with relatively smaller bias and RMSE, while there are a few instances where MML-EM produces estimates closer to the true value, for instance, the estimate of b_5 when the data is generated under $\mathcal{C} + SLM$. Nonetheless, most of these cases appear to fall within the uncertainty range expressed by the RMSE.

Table 3. Parameter estimation of KST-IRT models under fully misspecified structure \mathcal{K} (data-generating (DG) model and fitted model (FM) and state response probability $\pi(K|\theta)$ with Gibbs sampling estimation.

| ξ_j | $\pi(K \theta)$ | | \mathcal{K} | | EM-with-Gibbs | | | | |
|---------------|-----------------|-----|---------------|---------------|---------------|--------|--------|--------|--------|
| | | | | | j | | | | |
| | | | | | DG | FM | DG | FM | 1 |
| b_j | SLM | LKS | 2^Q | \mathcal{C} | −1.94 | −1.09 | 0.02 | 1.08 | 1.84 |
| | | | \mathcal{C} | 2^Q | (0.16) | (0.11) | (0.02) | (0.07) | (0.13) |
| | LKS | SLM | 2^Q | \mathcal{C} | −2.13 | −1.19 | 0.05 | 0.91 | 2.06 |
| | | | \mathcal{C} | 2^Q | (0.15) | (0.12) | (0.05) | (0.09) | (0.08) |
| η_{q_j} | SLM | LKS | 2^Q | \mathcal{C} | −1.90 | −1.12 | 0.05 | 1.08 | 1.85 |
| | | | \mathcal{C} | 2^Q | (0.12) | (0.14) | (0.04) | (0.11) | (0.13) |
| | LKS | SLM | 2^Q | \mathcal{C} | −2.31 | −1.16 | 0.06 | 0.96 | 2.10 |
| | | | \mathcal{C} | 2^Q | (0.27) | (0.14) | (0.08) | (0.11) | (0.09) |
| β_{q_j} | SLM | LKS | 2^Q | \mathcal{C} | 0.14 | 0.11 | 0.06 | 0.12 | 0.09 |
| | | | \mathcal{C} | 2^Q | (0.02) | (0.02) | (0.01) | (0.01) | (0.01) |
| | LKS | SLM | 2^Q | \mathcal{C} | 0.14 | 0.13 | 0.05 | 0.09 | 0.09 |
| | | | \mathcal{C} | 2^Q | (0.02) | (0.03) | (0.01) | (0.01) | (0.03) |
| β_{q_j} | SLM | LKS | 2^Q | \mathcal{C} | 0.13 | 0.09 | 0.09 | 0.12 | 0.10 |
| | | | \mathcal{C} | 2^Q | (0.02) | (0.01) | (0.02) | (0.03) | (0.04) |
| | LKS | SLM | 2^Q | \mathcal{C} | 0.14 | 0.11 | 0.05 | 0.11 | 0.10 |
| | | | \mathcal{C} | 2^Q | (0.03) | (0.01) | (0.02) | (0.01) | (0.02) |
| β_{q_j} | SLM | LKS | 2^Q | \mathcal{C} | 0.16 | 0.13 | 0.06 | 0.15 | 0.14 |
| | | | \mathcal{C} | 2^Q | (0.02) | (0.02) | (0.02) | (0.01) | (0.03) |
| | LKS | SLM | 2^Q | \mathcal{C} | 0.15 | 0.13 | 0.08 | 0.16 | 0.17 |
| | | | \mathcal{C} | 2^Q | (0.01) | (0.02) | (0.02) | (0.05) | (0.02) |
| β_{q_j} | SLM | LKS | 2^Q | \mathcal{C} | 0.16 | 0.10 | 0.05 | 0.14 | 0.15 |
| | | | \mathcal{C} | 2^Q | (0.02) | (0.01) | (0.01) | (0.02) | (0.01) |
| | LKS | SLM | 2^Q | \mathcal{C} | 0.15 | 0.10 | 0.05 | 0.13 | 0.15 |
| | | | \mathcal{C} | 2^Q | (0.02) | (0.02) | (0.01) | (0.03) | (0.03) |

Note. Each cell contains the average of the simulated estimates and the RMSE of the estimates in the parenthesis. b_j , η_{q_j} , and β_{q_j} represent the difficulty, lucky guess, and careless error parameters for item q_j .

It is noteworthy to observe that response patterns generated from a graded chain \mathcal{C} and fitted with a power set 2^Q appear to inflate the range of the difficulty parameters in all four designs, whereas response patterns generated from a power 2^Q and fitted with a graded chain \mathcal{C} appear to shrink the range of the difficulty parameters. This pattern only shows with the difficulty parameters but not with the conditional error parameters. This appears to be due to the structural role that the difficulty parameters play in the KST-IRT models. Indeed, since they determine the state response probabilities, they also establish whether these will be inflated or deflated. Contrary to the conditional error parameters, which simply cause low-level noise that reshuffles data among the different patterns of responses, difficulties have more leverage in shaping the values of the state probabilities and, therefore, of the final probability of a response pattern. This reflects variations in the range of the difficulty parameters when the data-generating structure shows discrepancies from the data-fitting one. Specifically, data generated under a graded chain would tend to have higher response probabilities for the Guttman patterns (as they come from higher values of the state probabilities of these patterns) than for the non-Guttman ones (as they only come from the noise generated by the conditional error parameters). As soon as such data is fitted by a power set, which instead expects to find higher values of the response probabilities for the non-Guttman patterns, the range of the difficulties is inflated in order to make the easiest items even easier, and the most difficult items even more difficult. This amounts to reducing the state response probabilities of the non-Guttman patterns, which the power set believes exist but are only very unlikely, in favor of the Guttman ones.

On the converse, data generated under a power set would tend to have more comparable values of the response probabilities for both Guttman and non-Guttman patterns. As soon as such data is fitted by a graded chain, which instead expects to find lower values of response probabilities of the non-Guttman patterns that are only due to noise generated by the conditional error parameters, the range of the difficulties is deflated in order to make the easiest items more difficult and the most difficult ones easier. This amounts to deflating the state response probabilities of the Guttman patterns, which according to the graded chains, should be the only existing ones, in favor of the non-Guttman ones so that the latter can have comparable response probabilities when data is reshuffled by the error parameters.

4.2. Simulation Study 2

The objective of this study is to compare the fit of the model between the RM and the KST-IRT models under misspecifications. The data-generating process and the fitted models are interchanged between RM and the KST-IRT models with two knowledge structures, \mathcal{C} and 2^Q , and two state response functions, SLM and LKS. As previously mentioned, however, the case $2^Q + LKS$ has been left out since it is formally equivalent to the case $2^Q + SLM$.

Similarly to the previous simulation study, for $N = 10,000$, response patterns were generated with the true parameter values from the previous simulation study. For the difficulty parameters, it was set to a vector of increasing difficulty parameters, $b = (-2, -1, 0, 1, 2)$, while the guessing and slipping parameters were set to the values $\eta = (0.129, 0.100, 0.074, 0.115, 0.096)$ and $\beta = (0.158, 0.112, 0.063, 0.148, 0.147)$. It should be noted that the KST-IRT models use all three parameters while the RM uses only the difficulty parameter. As highlighted in Section 2, this stresses that the two families of models assume very different data-generating processes.

For the computational method, though EM-with-Gibbs sampling outperforms MML-EM, the MML-EM is chosen as a method of parameter estimation as there are fewer assumptions involved. We use the MML-EM to estimate the parameters of the KST-IRT model and *rasch* function from *ltm* (version 1.2-0) package in R to estimate the RM difficulty parameters. Each MML-EM consumes, on average, 85.3 s to converge.

In Table 4, the first two columns indicate the data-generating (DG) model and fitted model (FM), respectively. As expected, the increasing trend of the difficulty parameters is captured in all six comparisons. When the KST-IRT models are fitted to RM-generated data, the estimates are closer to the true value than in the converse case. The fit of KST-IRT models onto the RM-generated data provided estimates that display consistent trends with only small differences. The highest absolute difference between the mean estimates of the KST-IRT model fit and the true value is 0.21, whereas that of the RM is as high as 1.10 at $j = 5$ for fitting RM to $\mathcal{C} + SLM$. However, the stability of the bias varies significantly when RM is used to fit the responses generated from KST-IRT models. As to the magnitudes of RMSEs, these are quite consistent across the model fit of the KST-IRT models, whereas fitting the RM to the KST-IRT responses is inconsistent across the items. For instance, the RMSE of fitting RM increases gradually to 1.11 when RM is fitted with $\mathcal{C} + SLM$. As a consequence, there appears to be two major concerns with the estimates of fitting RM to the assumed structure of the graded chain. Firstly, most of the estimates are overestimated compared to the true values. Secondly, the RMSE quickly rises as j increases. It does appear then that KST-IRT models, by virtue of their higher number of parameters, are more robust to violations of the data-generating process, while the traditional IRT approach, which assumes RM plus LI is more susceptible to violations of the LI assumption as it is implied by a data-generating process of the KST-IRT type in which a graded chain is considered. Notice that, instead, when the data is generated using a power set, the RM is less biased in its estimates since the combination of SLM and power set is equivalent to a 4PL model of IRT (with the discrimination parameter set to 1). Hence, the data generated from $2^Q + SLM$

can be considered as generated by a Rasch model with left-side added guessing and ceiling parameter, and where local independence is assumed.

Table 4. Difficulty parameter estimation of Rasch and KST-IRT models under misspecified data-generating (DG) model and fitted model (FM).

| DG | FM | b_j | | | | |
|-------------|-------------|-----------------|-----------------|-----------------|----------------|----------------|
| | | j | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
| RM | $C + SLM$ | −1.86 (0.38) | −0.92 (0.30) | −0.02 (0.23) | 0.93 (0.29) | 1.80 (0.46) |
| RM | $C + LKS$ | −1.89 (0.28) | −0.95 (0.36) | −0.00 (0.30) | 0.95 (0.41) | 1.87 (0.50) |
| RM | $2^Q + SLM$ | −1.79 (0.41) | −1.07 (0.37) | −0.04 (0.31) | 0.90 (0.34) | 1.87 (0.36) |
| $C + SLM$ | RM | −1.62 (0.38) | −0.52 (0.48) | 0.86 (0.86) | 2.04 (1.04) | 3.10 (1.11) |
| $C + LKS$ | RM | −1.90 (0.11) | −1.22 (0.22) | −0.00 (0.02) | 1.31 (0.31) | 2.45 (0.48) |
| $2^Q + SLM$ | RM | −1.60 (0.45) | −1.05 (0.14) | −0.05 (0.07) | 1.13 (0.20) | 2.05 (0.30) |

Note. Each cell contains the average of the simulated estimates and the RMSE of the estimates in the parenthesis. b_j , η_{q_j} , and β_{q_j} represent the difficulty, lucky guess, and careless error parameters for item q_j .

5. Discussion

The present contribution had a two-fold aim. On the one hand, to test and compare estimation methods for KST-IRT models. On the other hand, to compare the performance of these models to the more traditional IRT approach of using the Rasch model in conjunction with the assumption of LI to model a probabilistic version of a Guttman scale. This second question can be considered part of a larger investigation on the possible benefits of using a KST-IRT approach to model LD as suggested by Noventa et al. [18].

As to the first aim, two computational methods, MML-EM and EM-with-Gibbs, were presented. The steps to derive the marginalized likelihood of how closely the KST-IRT models follow the guidelines given by Bock and Lieberman [25] and Bock and Aitkin [26]. A noteworthy difference in using the KST-IRT models for the MML approach is that the joint likelihood of the response pattern does not assume local independence among items, so the joint likelihood does not take the form of a product of factorized item response functions. This difference, however, does not hinder the use of the MML-EM. The artificial measurements such as \bar{n}_k and \bar{r}_{jk} depend on the observed responses, not necessarily the likelihood built under LI. For the Gibbs sampling approach, selecting the proper prior distribution for the KST-IRT item parameters showed that the response data can update the posterior distributions, yielding comparable estimates to those of MML-EM.

The parameter estimates were consistent with the correct model specification, and the estimates deviated while capturing the trend when misspecified models were fit. EM-with-Gibbs sampling outperforms MML-EM overall by small margins. However, it must be noted that implementing Gibbs sampling involves more assumptions on the item parameters. The MML-EM also is not completely frequentist in that matter, and the estimates can suffer deviating estimates when the prior distribution of the ability parameter is misspecified. However, with the Gibbs sampling, as the posterior distributions are only updated accordingly to the predetermined distributions, misspecification can lead to undesirable results entirely.

The computational burden for the presented method increases with the incremental number of items, mainly due to the growing complexity of knowledge structures and the increasing number of item parameters associated with each knowledge state. Efficiently implementing the MML-EM with larger assessments necessitates obtaining the marginalized

likelihood and its essential components. For instance, obtaining derivatives of the item response functions with respect to each item parameter can bear a more computational burden, particularly with arbitrary knowledge structures beyond the graded chain and the power set. There is a need to generalize the parameter estimation method to handle larger assessments and accommodate arbitrary knowledge structures. This expansion, in the future, will offer a broader range of options for modeling responses, especially when the knowledge structure is capable of specifying the dependence structure.

Additionally, for further practical use, it is crucial to have model diagnostics to evaluate the model fit between the variations of the KST-IRT models. Beyond achieving reliable estimates with or without misspecification, model diagnostics to identify the better fit of the knowledge structure or the state response probability. Investigating model diagnostics will be essential in ensuring the appropriateness of the chosen model and enhancing the overall quality of the parameter process.

As to the second aim, it appears that KST-IRT models perform well in modeling a probabilistic version of a Guttman scale by capturing the essence of the scalogram by means of a graded chain, and by implementing additional noise in the form of conditional error parameters, to generate a non-Guttman pattern of responses. As such, they appear to provide an alternative to the more traditional IRT modeling of a probabilistic Guttman scale achieved by means of Rasch models under the assumption of local independence. As the two approaches are radically different in their fundamental assumptions, they were tested in terms of bias and robustness under the interchange of their respective generating processes. Generally, KST-IRT models appear to be less biased and more robust to violations of the data-generating process itself than the IRT traditional counterpart. Nonetheless, without having the true form of the Guttman scale or clear guidance on modeling the local dependence with the Rasch model, the direct comparison with the KST-IRT models that can inherently incorporate the local dependence and also allow for a larger number of parameters may not be fair. Nonetheless, for future consideration of more sophisticated IRT models and further acknowledging their prevalence in practice, a better understanding of the role of the noise parameters such as guessing and careless error probabilities, which in IRT applications are often debated or even neglected and yet in a KST-IRT approach are instead needed to recover those pattern of responses that do not conform with the knowledge structure, appears to be vital.

Author Contributions: Conceptualization, S.N. and A.K.; Methodology, S.Y.; Investigation, S.Y. and S.N.; Writing—original draft, S.Y. and S.N.; Writing—review and editing, S.N. and S.Y.; Supervision, S.N. and A.K.; Project administration, S.N. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Deutsche Forschungsgemeinschaft (DFG) Grant No. NO 1505/2-1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during the current study can be made available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---------|---|
| RM | Rasch Model |
| KST-IRT | Knowledge Space Theory—Item Response Theory |
| BLIM | Basic Local Independence Model |
| MML | Marginal Maximum Likelihood |
| EM | Expectation Maximization |

References

1. Lord, F.; Novik, M.R. *Statistical Theories of Mental Test Scores*; Addison-Wesley Publishing Company: London, UK, 1968.
2. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Nielsen & Lydiche: Copenhagen, Denmark, 1960.
3. Fischer, G.; Molenaar, I.W. *Rasch Models: Foundations, Recent Developments, and Applications*; Springer: New York, NY, USA, 1995.
4. Mokken, R. *A Theory and Procedure of Scale Analysis: With Applications in Political Research*; Walter de Gruyter, Mouton: New York, NY, USA; Berlin, Germany, 1971.
5. Scheiblechner, H. Isotonic ordinal probabilistic models. *Psychometrika* **1995**, *60*, 281–304. [[CrossRef](#)]
6. Sijtsma, K. Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Psychometrika* **1998**, *22*, 3–31. [[CrossRef](#)]
7. Yen, W.M. Scaling performance assessments: Strategies for managing local item dependence. *J. Educ. Meas.* **1993**, *30*, 187–213. [[CrossRef](#)]
8. Chen, W.H.; Thissen, D. Local dependence indexes for item pairs using item response theory. *J. Educ. Behav. Stat.* **1997**, *22*, 265–289. [[CrossRef](#)]
9. Falmagne, J.C.; Doignon, J.P. *Learning Spaces*; Springer: Berlin/Heidelberg, Germany, 2011.
10. Guttman, L. *The Basis for Scalogram Analysis: Measurement and Prediction*; John Wiley Sons: New York, NY, USA, 1950; Volume 4, pp. 60–90.
11. Mislevy, R.J.; Chang, H.H. Does adaptive testing violate local independence? *Psychometrika* **2000**, *65*, 149–156. [[CrossRef](#)]
12. Engelhard, G., Jr. Historical perspectives on invariant measurement: Guttman, Rasch, and Mokken. *Measurement* **2008**, *6*, 155–189. [[CrossRef](#)]
13. Andrich, D. An elaboration of Guttman scaling with Rasch models for measurement. *Sociol. Methodol.* **1985**, *15*, 33–80. [[CrossRef](#)]
14. Brink, N.E. Rasch's logistic model vs. the Guttman model. *Educ. Psychol. Meas.* **1972**, *32*, 921–927. [[CrossRef](#)]
15. Adams, R.J.; Wu, M.L.; Wilson, M. The Rasch rating model and the disordered threshold controversy. *Educ. Psychol. Meas.* **2012**, *72*, 547–573. [[CrossRef](#)]
16. Andrich, D. An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any “threshold disorder controversy”. *Educ. Psychol. Meas.* **2013**, *73*, 78–124. [[CrossRef](#)]
17. Tutz, G. On the structure of ordered latent trait models. *J. Math. Psychol.* **2020**, *96*, 102346. [[CrossRef](#)]
18. Noventa, S.; Spoto, A.; Heller, J.; Kelava, A. On a generalization of local independence in item response theory based on knowledge space theory. *Psychometrika* **2019**, *84*, 395–421. [[CrossRef](#)] [[PubMed](#)]
19. van der Linden, W.J. *Handbook of Item Response Theory: Three Volume Set*; CRC Press: Boca Raton, FL, USA, 2018.
20. Barton, M.; Lord, F. *An Upper Asymptote for the Three-Parameter Logistic Item-Response Model*; Research Report RR-81-20; Educational Testing Service: Princeton, NJ, USA, 1981.
21. Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*; Lord, F.M., Novick, M.R., Eds.; Addison-Wesley: Boston, MA, USA, 1968; Chapter 17–20.
22. Heller, J.; Stefanutti, L.; Anselmi, P.; Robusto, E. On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika* **2015**, *80*, 995–1019. [[CrossRef](#)] [[PubMed](#)]
23. Noventa, S.; Heller, J.; Stefanutti, L. Some considerations on the factorization of state probabilities in knowledge structures. *J. Math. Psychol.* **2021**, *102*, 102542. [[CrossRef](#)]
24. Stefanutti, L. A logistic approach to knowledge structures. *J. Math. Psychol.* **2006**, *50*, 545–561. [[CrossRef](#)]
25. Bock, D.; Lieberman, M. Fitting a response model for n dichotomously scored items. *Psychometrika* **1970**, *35*, 179–197. [[CrossRef](#)]
26. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459.
27. Albert, J.H. Bayesian estimation of normal ogive item response curves using Gibbs sampling. *J. Educ. Stat.* **1992**, *17*, 251–269.
28. Engelen, R. *A Review of Different Estimation Procedures in the Rasch Model*; Number 87-6 in OMD research report; Project Psychometric Aspects of Item Banking No. 21; Faculty Educational Science and Technology, University of Twente: Enschede, The Netherlands, 1987. [[CrossRef](#)]
29. Harwell, M.R.; Baker, F.B. The use of prior distributions in marginalized Bayesian item parameter estimation: A didactic. *Appl. Psychol. Meas.* **1991**, *15*, 375–389. [[CrossRef](#)]
30. Mislevy, R.J. Bayes modal estimation in item response models. *Psychometrika* **1986**, *51*, 177–195. [[CrossRef](#)]
31. Bock, R.D.; Mislevy, R.J. Adaptive EAP estimation of ability in a microcomputer environment. *Appl. Psychol. Meas.* **1982**, *6*, 431–444. [[CrossRef](#)]
32. Cai, L. High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika* **2010**, *75*, 33–57. [[CrossRef](#)]
33. Cai, L. Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *J. Educ. Behav. Stat.* **2010**, *35*, 307–335. [[CrossRef](#)]
34. Patz, R.J.; Junker, B.W. A straightforward approach to Markov chain Monte Carlo methods for item response models. *J. Educ. Behav. Stat.* **1999**, *24*, 146–178. [[CrossRef](#)]

35. Cowles, M.K. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Stat. Comput.* **1996**, *6*, 101–111.
36. Hambleton, R.K.; Swaminathan, H.; Rogers, H.J. *Fundamentals of Item Response Theory*; Sage: New York, NY, USA, 1991; Volume 2.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.