Spanish Journal of Psychology (2015), **18**, e26, 1–7.

doi:10.1017/sjp.2015.24

CAMBRIDGE UNIVERSITY PRESS

Naïve Tests of Basic Local Independence Model's Invariance

Debora de Chiusole, Luca Stefanutti, Pasquale Anselmi and Egidio Robusto

Università di Padova (Italy)

Abstract. The *basic local independence model* (BLIM) is a probabilistic model for knowledge structures, characterized by the property that lucky guess and careless error parameters of the items are independent of the knowledge states of the subjects. When fitting the BLIM to empirical data, a good fit can be obtained even when the invariance assumption is violated. Therefore, statistical tests are needed for detecting violations of this specific assumption. This work provides an extension to theoretical results obtained by de Chiusole, Stefanutti, Anselmi, and Robusto (2013), showing that statistical tests based on the partitioning of the empirical data set into two (or more) groups are not adequate for testing the BLIM's invariance assumption. A simulation study confirms the theoretical results.

Received 2 July 2013; Revised 22 May 2014; Accepted 3 October 2014

Keywords: knowledge space theory, basic local independence model, parameter invariance assumption, bipartition models.

Knowledge space theory (KST) is a mathematical theory, developed by Doignon and Falmagne (1985, 1999) and Falmagne and Doignon (2011) for the individual knowledge assessment. The aim of a KST assessment is to find, with maximum accuracy and efficiency, the set of problems that a student masters in a specific knowledge domain (the student's *knowledge state*). In real contexts, some of the student's answers could be affected by noise, like careless errors due to time pressure or to turmoil, or correct answers resulting from guessing. Because of the noise, the student's knowledge state is not directly observable, and it has to be inferred from the student's responses to the problems. To provide realistic predictions of student's responses, probabilistic models have to be considered. The first and the most used probabilistic model, developed in KST, is the so-called basic local independence model (BLIM; Falmagne & Doignon (1988a, 1988b).

Knowledge about the properties of this model has grown over the years. For example, methods for estimating its parameters (Heller & Wickelmaier, 2013; Schrepp, 2005; Stefanutti & Robusto, 2009) and for testing its identifiability (Spoto, Stefanutti, & Vidotto, 2012; Stefanutti, Heller, Anselmi, & Robusto, 2012) are available. Furthermore, some extensions of the BLIM have been proposed, like, for example, the Gain-Loss Model (Robusto, Stefanutti, & Anselmi, 2010; de Chiusole, Anselmi, Stefanutti, & Robusto, 2013), a model for assessing learning processes, and a probabilistic model for skill dependence (de Chiusole & Stefanutti, 2013) were developed and applied to real data.

The focus of this article is on one of the properties of the BLIM, that is the *parameter invariance assumption*. In de Chiusole, Anselmi, et al. (2013) de Chiusole, Stefanutti et al. (2013) it is shown that, even when the invariance assumption is violated by the data, the goodness of fit of the BLIM might be acceptable. For this reason, having a method that shows up invariance violations becomes essential. The method proposed by the authors consists in comparing the BLIM with other models, called bipartition models (BPMs), in which the invariance assumption is explicitly violated. If the comparison favors a BPM, then the parameter invariance assumption is violated, meaning that the BLIM is not adequate for those data.

In the same article, another method to discover such type of violations was considered. The method was inspired by the IRT approach (Andersen, 1973; Glas & Verhelst, 1995), and consists in partitioning the observed data set into two or more independent groups, to fit the BLIM in each of the groups, and to apply some suitable statistical test to evaluate the difference between the parameter estimates of the two groups. In the simplest case, two groups are formed by separating all the subjects with score levels below the median from those with score levels above the median. If the test is significant, then the conclusion would be that the parameter invariance assumption is violated.

Even if this method seems to be the most natural way to discover violations, it has been formally proven that, with the BLIM, it does not work properly. The main concern with this method is that the error parameter estimates in the two groups will significantly

Correspondence concerning this article should be addressed to Debora de Chiusole. Università di Padova. FISPPA Department. Via Venenzia, 12. 35131. Padua (Italy).

E- mail: deboratn@libero.it

differ one another even when the invariance assumption is, indeed, respected. For this reason, in the sequel we refer to this method as the *naïve test of invariance*.

A procedure, similar to the one described above, was developed by de La Torre and Lee (2010) in the area of cognitive diagnostic models, for showing up violations of the parameter invariance of the DINA (deterministic inputs, noisy AND-gate) model (Junker & Sijtsma, 2001). Instead of forming two 'pure' groups below and above the median, they produced two data sets that were mixtures of the two pure groups. A first group collected about 60% of respondents that were below the median, and about 40% of those that were above it. The second group was constructed by reversing these proportions. By applying that procedure to a data set on fraction subtraction, de la Torre and Lee concluded that the parameter invariance of the DINA model may not hold in real data. It is worth noting that, on the performance level, the DINA model is equivalent to the BLIM (Heller, Stefanutti, Anselmi, & Robusto, 2014).

The aim of this article is to generalize the theoretical results concerning inadequacy of the naïve test to any choice of the proportion p, used to form the two groups. After presenting the BLIM, naïve tests of invariance are presented along with theoretical results showing that also the general version of the test suffers for the same problems. The theoretical results are illustrated through a simulation study and an empirical application.

The BLIM and the Parameter Invariance

The aim of a KST assessment is to uncover the knowledge state that characterizes a student, on the basis of her responses to a given set Q of problems. The collection of responses is named *response pattern*, and it is represented by the subset $R \subseteq Q$ of all problems that received a correct response. For a KST assessment, a deterministic model on all problems $q \in Q$, called knowledge structure, is required, along with a probabilistic model, like for example, the BLIM. A knowledge structure is defined as a pair $(Q, \mathbf{K})$, in which Q is a collection of problems, and $\mathbf{K}$ is a collection of subsets of Q, called *knowledge states*. The BLIM is defined as a quadruple $(Q, \mathbf{K}, \pi, r)$, in which:

a. (Q, K) is a knowledge structure on a finite set Q;
b. π is a probability distribution on $\mathbf{K}$
c. r is the *response function* and, for every $R \subseteq Q$ and every $K \in \mathbf{K}$, it is defined by

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

where β_q, $\eta_q \in (0, 1]$ are two parameters of each of the items, respectively called *careless error* probability and *lucky guess* probability;

The probability of sampling a student whose response pattern is $R \subseteq Q$ is

$$P(R) = \sum_{K \in \mathbf{K}} r(R, K)\pi_K. \quad (2)$$

It has to be pointed out that the parameters β_q and η_q are attached to the items and do not vary with the knowledge states of the students, in other words they are invariant across individuals. We will refer to this property as the *invariance assumption*.

Before to get into the question of how this assumption can be tested, it is worth providing some basic notation. The set $\mathbf{K}_q = \{K \in \mathbf{K} : q \in K\}$ collects all knowledge states containing a given item $q \in Q$, and $\bar{\mathbf{K}}_q = \{K \in \mathbf{K} : q \notin K\}$ is its complement in $\mathbf{K}$. Similarly, for $\mathbf{R} = 2^Q$, let $\mathbf{R}_q = \{R \subseteq Q : q \in R\}$ be the set of all response patterns containing q, and $\bar{\mathbf{R}}_q = \{R \subseteq Q : q \notin R\}$ be its complement in $\mathbf{R}$. Finally, for any $\mathbf{F} \subseteq \mathbf{R}$ and any $\mathbf{J} \subseteq \mathbf{K}$ let

$$P(\mathbf{F}, \mathbf{J}) = \sum_{R \in \mathbf{F}} \sum_{K \in \mathbf{J}} r(R, K)\pi_K$$

be the joint probability of $\mathbf{F}$ and $\mathbf{J}$.

Naïve Tests of Invariance: Restricted Case

A way to assess violations of the parameter invariance assumption of the BLIM would be to partition the whole data set into two independent groups, to fit the BLIM in each of them (say, Group 1 and Group 2), and to apply some suitable statistical test of the difference between the parameter estimates in the two groups. If the test is significant then the conclusion would be that the parameter invariance is violated by the data.

As a criterion to form the two groups, consider the one that consists of choosing a certain quantile $c > 0$ (e.g. the median) of the sample distribution of the size of the response patterns. Those having size less or equal to c are assigned to Group 1, and those having size greater than c are assigned to Group 2. With these two groups, a test of the invariance would not work properly, since parameter estimates would be biased in both groups, even when the independence assumption is, indeed, respected. Because of this bias, the statistical test would lead to a rejection of the local independence assumption too often.

To see this, consider some cutoff $c \in \{0, 1, \ldots, |Q| - 1\}$, and let $\mathbf{R}^{\downarrow} = \{R \in \mathbf{R} : |R| \leq c\}$ be the collection of all response patterns whose size is less or equal to c, and $\mathbf{R}^{\uparrow} = \{R \in \mathbf{R} : |R| > c\}$ be the collection of all response patterns whose size is greater than c. Then, according to the BLIM, the conditional probability that in a randomly sampled response pattern R, an item q is

https://doi.org/10.1017/sjp.2015.24 Published online by Cambridge University Press

failed by careless error, given that the size of R is below the cutoff is

$$\beta_q^{\downarrow} = \frac{\sum_{R \in \bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}} \sum_{K \in \mathbf{K}_q} r(R,K)\pi_K}{\sum_{R \in \mathbf{R}^{\downarrow}} \sum_{K \in \mathbf{K}_q} r(R,K)\pi_K}, \quad (3)$$

whereas the conditional probability that in a randomly sampled response pattern R, item q is solved by lucky guess, given that R is below the cutoff is

$$\eta_q^{\downarrow} = \frac{\sum_{R \in \mathbf{R}_q \cap \mathbf{R}^{\downarrow}} \sum_{K \in \bar{\mathbf{K}}_q} r(R,K)\pi_K}{\sum_{R \in \mathbf{R}^{\downarrow}} \sum_{K \in \bar{\mathbf{K}}_q} r(R,K)\pi_K}, \quad (4)$$

Similar equations are obtained for the $\beta_q^{\uparrow}$ and $\eta_q^{\uparrow}$ parameters, by replacing $\mathbf{R}^{\downarrow}$ with $\mathbf{R}^{\uparrow}$ in Equations 3 and 4.

In de Chiusole, Anselmi et al. (2013), it is shown that, for any choice of the cutoff c and any item $q \in Q$ and $\beta_q, \eta_q \in (0, 1)$, the following inequalities hold true: $\beta_q^{\uparrow} < \beta_q < \beta_q^{\downarrow}$; and $\eta_q^{\downarrow} < \eta_q < \eta_q^{\uparrow}$.

These two inequalities show that careless errors are more likely when one samples below the cutoff, whereas lucky guesses are more likely when one samples above the cutoff. Thus, when the parameters of the BLIM are estimated from only a part of the data set (below/above), one obtains biased parameter estimates.

Naïve Tests of Invariance: General Case

It might be argued that the rule of forming the two groups by assigning all patterns below a certain cutoff to one group, and all the remaining ones to the other group is too strong. Maybe, there exist weaker rules that reduce or even remove the bias.

The following, more general, rule is considered here: Given some proportion $p^{\downarrow}$, with $0 \le p^{\downarrow} \le 1$, a sufficiently large number $n^{\downarrow}$ of response patterns are randomly sampled with replacement from $\mathbf{R}^{\downarrow}$, and each of them is assigned to Group $\mathbf{G}_1$ with probability $p^{\downarrow}$, and to Group $\mathbf{G}_2$ with probability $1 - p^{\downarrow}$. Analogously, given a proportion $p^{\uparrow}$, $0 \le p^{\uparrow} \le 1$, $n^{\uparrow}$ response patterns are randomly sampled with replacement from $\mathbf{R}^{\uparrow}$, and each of them is assigned to $\mathbf{G}_1$ with probability $p^{\uparrow}$, and to $\mathbf{G}_2$ with probability $1 - p^{\uparrow}$.

As the following proposition shows, even the more general rule suffers from the same problem as the naïve test of invariance.

Proposition 1

Let $\beta_q^{(1)}$ be the probability that, in a randomly sampled response pattern, an item q is failed by careless error, given that the pattern belongs to Group $\mathbf{G}_1$, and $\eta_q^{(1)}$ be the probability that a lucky guess occurs for q, given that the pattern belongs to $\mathbf{G}_1$. Then $\beta_q^{(1)} \le \beta_q$ if and only if

$$\frac{p^{\downarrow}}{p^{\uparrow}} \le \frac{\beta_q P(\mathbf{R}^{\uparrow},\mathbf{K}_q) - P(\bar{\mathbf{R}}_q^{\uparrow},\mathbf{K}_q)}{P(\bar{\mathbf{R}}_q^{\downarrow},\mathbf{K}_q) - \beta_q P(\mathbf{R}^{\downarrow},\mathbf{K}_q)}, \quad (5)$$

where $\bar{\mathbf{R}}_q^{\downarrow} = \bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}$ and $\bar{\mathbf{R}}_q^{\uparrow} = \bar{\mathbf{R}}_q \cap \mathbf{R}^{\uparrow}$. Moreover, $\eta_q^{(1)} \le \eta_q$ if and only if

$$\frac{p^{\downarrow}}{p^{\uparrow}} \le \frac{\eta_q P(\mathbf{R}^{\uparrow},\bar{\mathbf{K}}_q) - P(\mathbf{R}_q^{\uparrow},\bar{\mathbf{K}}_q)}{P(\mathbf{R}_q^{\downarrow},\bar{\mathbf{K}}_q) - \eta_q P(\mathbf{R}^{\downarrow},\bar{\mathbf{K}}_q)}, \quad (6)$$

where $\mathbf{R}_q^{\downarrow} = \mathbf{R}_q \cap \mathbf{R}^{\downarrow}$ and $\mathbf{R}_q^{\uparrow} = \mathbf{R}_q \cap \mathbf{R}^{\uparrow}$.

What Proposition 1 essentially says is that the method that consists in partitioning the sample into two subgroups by using any arbitrary proportions $p^{\uparrow}$ and $p^{\downarrow}$ will lead to biased estimates of the β_q and η_q parameters, even when the invariance assumption is indeed respected by the data. Depending on the ratio $p^{\downarrow}/p^{\uparrow}$ that one chooses, the β_q and η_q probabilities might be either over- or underestimated in both groups $\mathbf{G}_1$ and $\mathbf{G}_2$. For this reason it is recommended not using methods like the one described in this section for testing parameter invariance of the BLIM.

Proof of Proposition 1

Suppose that the probabilities of the response patterns in a population are given by Equation (2). The probability that in a randomly sampled response pattern, an item q is failed by careless error, given that the pattern belongs to group $\mathbf{G}_1$ is

$$\beta_q^{(1)} = P(\bar{\mathbf{R}}_q \mid \mathbf{K}_q, \mathbf{G}_1) = \frac{P(\bar{\mathbf{R}}_q \cap \mathbf{G}_1, \mathbf{K}_q)}{P(\mathbf{K}_q, \mathbf{G}_1)}. \quad (7)$$

The numerator of the right hand side of Equation (7) can be written as $P(\bar{\mathbf{R}}_q \cap \mathbf{G}_1, \mathbf{K}_q) = P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow} \cap \mathbf{G}_1, \mathbf{K}_q) + P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\uparrow} \cap \mathbf{G}_1, \mathbf{K}_q)$, and, by applying the concatenation rule of conditional probabilities, $P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow} \cap \mathbf{G}_1, \mathbf{K}_q) = P(\mathbf{G}_1 \mid \bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}, \mathbf{K}_q) P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}, \mathbf{K}_q)$. Moreover, given $\mathbf{R}^{\downarrow}$, $\mathbf{G}_1$ is independent of both $\bar{\mathbf{R}}_q$ and $\mathbf{K}_q$, hence

$$P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow} \cap \mathbf{G}_1, \mathbf{K}_q) = P(\mathbf{G}_1 \mid \mathbf{R}^{\downarrow}) P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}, \mathbf{K}_q) = p^{\downarrow} P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}, \mathbf{K}_q).$$

Similarly, we have $P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\uparrow} \cap \mathbf{G}_1, \mathbf{K}_q) = P(\mathbf{G}_1 \mid \mathbf{R}^{\uparrow}) P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\uparrow}, \mathbf{K}_q) = p^{\uparrow} P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\uparrow}, \mathbf{K}_q)$. Therefore, the numerator of (7) becomes $P(\bar{\mathbf{R}}_q \cap \mathbf{G}_1, \mathbf{K}_q) = p^{\downarrow} P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}, \mathbf{K}_q) + p^{\uparrow} P(\bar{\mathbf{R}}_q \cap \mathbf{R}^{\uparrow}, \mathbf{K}_q)$. On the other hand, the denominator of (7) can be written as

$$\begin{aligned} P(\mathbf{K}_q, \mathbf{G}_1) &= P(\mathbf{K}_q, \mathbf{G}_1 \cap \mathbf{R}^{\downarrow}) + P(\mathbf{K}_q, \mathbf{G}_1 \cap \mathbf{R}^{\uparrow}) \\ &= P(\mathbf{G}_1 \mid \mathbf{R}^{\downarrow}, \mathbf{K}_q) P(\mathbf{R}^{\downarrow}, \mathbf{K}_q) + P(\mathbf{G}_1 \mid \mathbf{R}^{\uparrow}, \mathbf{K}_q) P(\mathbf{R}^{\uparrow}, \mathbf{K}_q) \\ &= p^{\downarrow} P(\mathbf{R}^{\downarrow}, \mathbf{K}_q) + p^{\uparrow} P(\mathbf{R}^{\uparrow}, \mathbf{K}_q). \end{aligned}$$

https://doi.org/10.1017/sjp.2015.24 Published online by Cambridge University Press

Thus, with the notation $\bar{\mathbf{R}}_q^{\downarrow} = \bar{\mathbf{R}}_q \cap \mathbf{R}^{\downarrow}$, and $\bar{\mathbf{R}}_q^{\uparrow} = \bar{\mathbf{R}}_q \cap \mathbf{R}^{\uparrow}$, Equation (7) can be rewritten as

$$\beta_q^{(1)} = \frac{p^{\downarrow} P(\bar{\mathbf{R}}_q^{\downarrow}, \mathbf{K}_q) + p^{\uparrow} P(\bar{\mathbf{R}}_q^{\uparrow}, \mathbf{K}_q)}{p^{\downarrow} P(\mathbf{R}^{\downarrow}, \mathbf{K}_q) + p^{\uparrow} P(\mathbf{R}^{\uparrow}, \mathbf{K}_q)}. \quad (8)$$

By substituting $\beta_q^{(1)}$ with the right hand of this last equation in the inequality $\beta_q^{(1)} \le \beta_q$, after some algebra we obtain

$$p^{\downarrow}[P(\bar{\mathbf{R}}_q^{\downarrow}, \mathbf{K}_q) - \beta_q P(\mathbf{R}^{\downarrow}, \mathbf{K}_q)] \le p^{\uparrow}[\beta_q P(\mathbf{R}^{\uparrow}, \mathbf{K}_q) - P(\bar{\mathbf{R}}_q^{\uparrow}, \mathbf{K}_q)]. \quad (9)$$

We know from Proposition 1 in de Chiusole, Anselmi, et al., 2013; de Chiusole & Stefanutti (2013) that $\beta_q > \beta_q^{\uparrow}$. Since, by definition, $\beta_q^{\uparrow} = P(\bar{\mathbf{R}}_q^{\uparrow}, \mathbf{K}_q)/P(\mathbf{R}^{\uparrow}, \mathbf{K}_q)$, we have that $\beta_q P(\mathbf{R}^{\uparrow}, \mathbf{K}_q) - P(\bar{\mathbf{R}}_q^{\uparrow}, \mathbf{K}_q) > 0$. Thus, dividing both terms of the inequality in (9) by $p^{\uparrow}$ and, then, by $P(\bar{\mathbf{R}}_q^{\downarrow}, \mathbf{K}_q) - \beta_q P(\mathbf{R}^{\downarrow}, \mathbf{K}_q)$, one obtains the Inequality in (5).

The proof concerning the conditions for $\eta_q \le \eta_q^{(1)}$ follows an identical line of reasoning, provided that $\bar{\mathbf{R}}_q$ is replaced by $\mathbf{R}_q$ and $\mathbf{K}_q$ by $\bar{\mathbf{K}}_q$.

Proposition 1 holds true with arbitrary values of the two proportions $p^{\downarrow}$ and $p^{\uparrow}$. A special case arises when the choice of $p^{\downarrow}$ and $p^{\uparrow}$ is such that $p^{\uparrow} = 1 - p^{\downarrow}$.

Proposition 2

If $p^{\uparrow} = 1 - p^{\downarrow}$ *then:*

1. $\beta_q^{(1)} \le \beta_q$ if and only if $\beta_q^{(2)} \ge \beta_q$;
2. $\eta_q^{(1)} \le \eta_q$ if and only if $\eta_q^{(2)} \ge \eta_q$.

Proof. (1) Let $f(\beta_q)$ represent the right hand term in the inequality (5). The we have $\beta_q^{(1)} \le \beta_q$ iff $p^{\downarrow}/(1-p^{\downarrow}) \le f(\beta_q)$ iff $(1-p^{\downarrow})/p^{\downarrow} \ge f(\beta_q)$ iff $p^{\uparrow}/(1-p^{\uparrow}) \ge f(\beta_q)$ iff $\beta_q^{(2)} \ge \beta_q$. An analogous development, applied to inequality (1), leads to condition (2).

As Equation (8) shows, the value of $\beta_q^{(1)}$ is a function of: (i) the cutoff c used to partition the dataset; (ii) the values of $p^{\uparrow}$ and $p^{\downarrow}$; (iii) the BLIM's parameter values (i.e., the β_q, η_q and π_K probabilities). To illustrate propositions 1 and 2, Figure 1 shows how $\beta_q^{(1)}$ varies as a function of the true parameter β_q and the proportion $p^{\downarrow}$. In the figure the x-axis represents the parameter β_q and each of the curves corresponds to a different choice of $p^{\downarrow}$. The remaining parameters of the BLIM were fixed to constant values, and the restriction $p^{\uparrow} = 1 - p^{\downarrow}$ was used.

It can be seen from the figure that, when $p^{\downarrow}$ is less than a certain value[1] (.5 in this particular example), the β_q parameter is underestimated in group $\mathbf{G}_1$ (and thus overestimated in group $\mathbf{G}_2$), and the size of the bias increases as $p^{\downarrow}$ approaches zero. On the other hand, when $p^{\downarrow}$ is greater than .5, the β_q parameter is overestimated in group $\mathbf{G}_1$ (and thus underestimated in group $\mathbf{G}_2$), and the size of the bias increases as $p^{\downarrow}$ approaches one. A similar example could be provided for the η_q parameter. In that case one obtains an opposite behavior.

[1]The value is obtained by replacing the inequality $\le$ in (5) with an equality and solving for $p^{\downarrow}$, with the constraint $p^{\downarrow} = 1 - p^{\uparrow}$.

A Simulation Study

The theoretical results obtained in previous section are illustrated by means of a simulation study in which the two proportions $p^{\downarrow}$ and $p^{\uparrow}$ are varied systematically. The aim of the simulations is to show that, by estimating the BLIM's parameters in each of the two groups $\mathbf{G}_1$ and $\mathbf{G}_2$, leads to reject the error parameter invariance assumption of the BLIM even when it is respected by the data, irrespectively of the values of the proportions $p^{\downarrow}$ and $p^{\uparrow}$. In particular, it is expected that on the average, the maximum likelihood estimates of the β_q and η_q parameters in group $\mathbf{G}_i$ (with $i \in \{1,2\}$) approach the theoretical values $\beta_q^{(i)} \ne \beta_q$ and $\eta_q^{(i)} \ne \eta_q$.

For all the simulations a set of MATLAB functions, available on request to the first author, were developed.

Simulation Design

A number of 9 simulation conditions were considered in which the following variables were held fixed: a random knowledge structure, composed of 16 items and 400 knowledge states[2]; the true error parameter values, chosen at random from a uniform distribution in the interval (0,.25); the cutoff used for creating the two groups, that, for this knowledge structure, was the median (8). What varied among the 9 conditions was the proportion $p^{\downarrow}$ used to form groups $\mathbf{G}_1$ and $\mathbf{G}_2$. The values of $p^{\downarrow}$ were taken from the open interval (0,1) at equally spaced intervals of length 0.10. In all simulations the constraint $p^{\uparrow} = 1 - p^{\downarrow}$ held true.

In each of the 9 conditions, 100 samples of 1,000 response patterns were generated. For each sample, the two groups were then formed choosing, with replacement, a proportion $p^{\downarrow}$ of the patterns below the median and a proportion $1-p^{\downarrow}$ of the patterns above the median. In this way, for each of the 100 replications, 9 pairs {**G**1, **G**2}, both composed of 1,000 response patterns, were obtained. The BLIM was then estimated to both groups, in each of the 9 pairs, and the means of the parameter estimates were compared to those computed by applying Equation (8) for $\beta_q^{(1)}$ and the corresponding equation for $\eta_q^{(1)}$

[2]The number of items/states of the structure was not extensively varied. Nonetheless this particular choice suffices as a counterexample showing that the naïve method does not work in general.

https://doi.org/10.1017/sjp.2015.24 Published online by Cambridge University Press

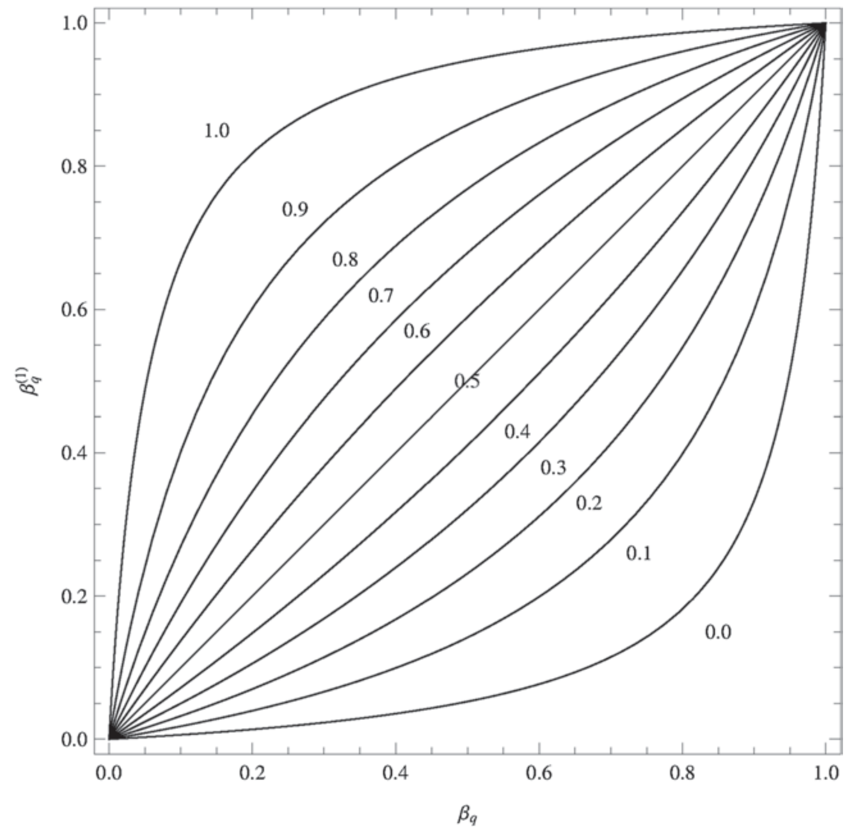

Figure 1. The probability $\beta_q^{(1)}$ of a careless error on item q in group G_1 (y-axis) varies as a function of β_q (x-axis). Each of the curves corresponds to a different choice of the proportion $p^{\downarrow}$.

Results

The size and the direction of the bias, in the 9 simulation conditions, were examined. Figure 2 shows the results for the β_q error parameter, obtained for both $\mathbf{G}_1$ and $\mathbf{G}_2$. In the figure, the true value of β_q is along the x-axis, whereas the theoretical and the mean estimates of β_q are along y-axis. The straight line is for reference and indicates that $x = y$, and there is one diagram for each of the 9 values of p. In each diagram, circles represent $\beta_q^{(1)}$, whereas triangles represent the theoretical value of $\beta_q^{(2)}$. Finally, the × and dots represent respectively the $\overline{\beta}_q^{(1)}$ and the $\overline{\beta}_q^{(2)}$ mean estimates.

From the figure, it can be seen that: (1) $\beta_q^{(i)}$ and $\overline{\beta}_q^{(i)}$ parameters are in agreement for both groups, in all 9 conditions; (2) when $p = .50$, no bias is observed; (3) going from $p = .50$ to $p = .10$, the β_q parameter is overestimated in $\mathbf{G}_2$ and underestimated in $\mathbf{G}_1$, whereas going from $p = .50$ to $p = .90$, the β_q parameter is underestimated in $\mathbf{G}_2$ and overestimated in $\mathbf{G}_1$. These results are in line with the predictions made by Proposition 2.

The results obtained for the η_q error parameters are very similar to those obtained for the β_q. The difference is that going from $p = .50$ to $p = .10$, the η_q parameter is underestimated in $\mathbf{G}_2$ and overestimated in $\mathbf{G}_1$, whereas going from $p = .50$ to $p = .90$, the η_q parameter is overestimated in $\mathbf{G}_2$ and underestimated in $\mathbf{G}_1$.

Empirical Application

The results discussed in the previous sections are illustrated by an application to real data. The design was the same used in the simulation study: 9 conditions where considered in which the proportion $p^{\downarrow}$ used to create the two groups **G**1 and **G**2, respectively below and above the cutoff c, varied in the open interval (0,1) at equally spaced interval of length .10. For illustrative purposes, the data set provided by de Chiusole, Anselmi, et al. (2013 de Chiusole, Stefanutti et al. (2013) was used, in which 18 problems of elementary probability theory (with a knowledge structure of 69 states) were administered to 209 Italian university students. The median of the cardinality of the knowledge states was used as the cutoff ($c = 9$) to form the two groups. Subsequently, in each of the 9 conditions, the BLIM was fitted to the data in each of the two groups below and above the cutoff, and the means of the β_q and η_q estimates were computed across the items, and compared to one another.

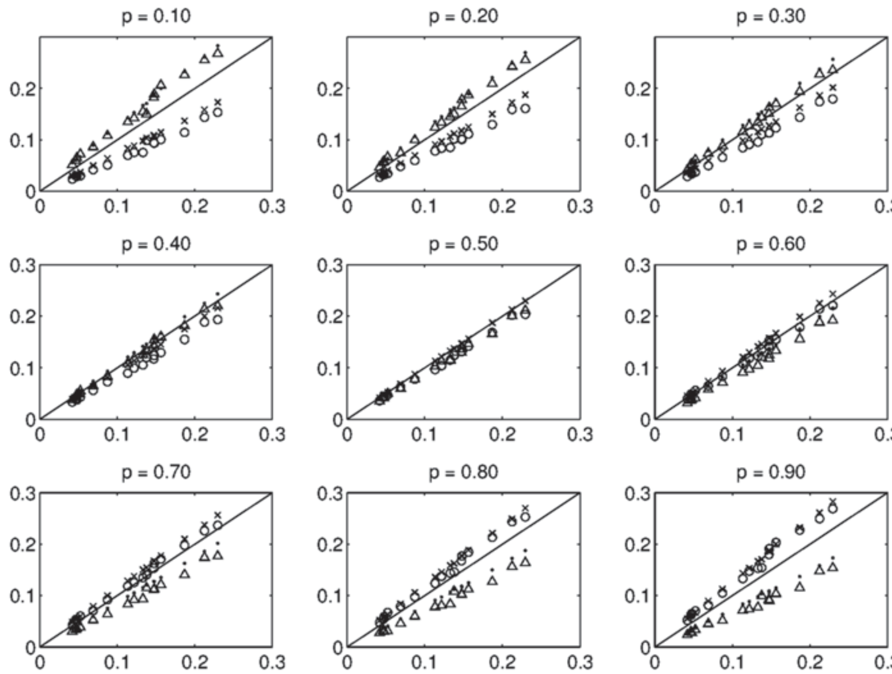

Figure 2. Comparison between $\beta_q^{(i)}$ and $\overline{\beta}_q^{(i)}$ parameters in $\mathbf{G}_1$ and $\mathbf{G}_2$. The true value of β_q is along the x-axis, whereas the theoretical and the mean estimates of β_q are along y-axis. The straight line is for reference and indicates that $x = y$, and there is one diagram for each of the 9 values of p. In each diagram, circles represent $\beta_q^{(1)}$, whereas triangles represent the theoretical value of $\beta_q^{(2)}$. Finally, the × and dots represent respectively the $\overline{\beta}_q^{(1)}$ and the $\overline{\beta}_q^{(2)}$ mean estimates.

Results

The results are shown in Table 1. Concerning the careless error parameters, it can be seen that in each of the conditions from 1 to 5, in which $.1 \le p^{\downarrow} < .5$, the inequality $\overline{\beta}_{G1} < \overline{\beta}_{G2}$ is respected, with the only exception of condition 4 (where, however, the two means are very close to one another); whereas, in all conditions from 6 to 9, in which $.5 \le p \le .9$, the inequality $\overline{\beta}_{G1} > \overline{\beta}_{G2}$ is respected. Concerning the lucky guess parameters, it can be seen that in all conditions from 1 to 5, the inequality $\overline{\eta}_{G1} > \overline{\eta}_{G2}$ holds true, whereas in all conditions from 6 to 9, the inequality $\overline{\eta}_{G1} < \overline{\eta}_{G2}$ holds. All these results are in line with Proposition 2.

The BLIM's parameter invariance says that the probability of a careless error or a lucky guess, for an item, does not depend on the student's knowledge state. In de Chiusole, Anselmi, et al. (2013), de Chiusole, Stefanutti et al. (2013) two methods for testing this assumption were presented and discussed. The former, consists in comparing the BLIM with other models, called bipartition models (BPMs), in which the invariance

Table 1. *Comparison among the mean of the item parameter estimates of the BLIM in the 9 conditions of the study. In the table* p *is the proportion used to create the two groups below and above the cutoff; the* $\overline{\beta}_{G1}$ *and* $\overline{\beta}_{G2}$ *parameters are the mean of the careless errors of the groups below and above the cutoff, respectively; the* $\overline{\eta}_{G1}$ *and* $\overline{\eta}_{G2}$ *parameters are the mean of the lucky guesses of the groups below and above the cutoff, respectively*

Condition	p	$\overline{\beta}_{G1}$	$\overline{\beta}_{G2}$	$\overline{\eta}_{G1}$	$\overline{\eta}_{G2}$
1	.10	.1523	.4195	.3519	.1335
2	.20	.2015	.3511	.3294	.1711
3	.30	.2113	.3439	.2617	.1722
4	.40	.3240	.3135	.2152	.2140
5	.50	.2401	.3400	.2316	.1715
6	.60	.2977	.2643	.1706	.1973
7	.70	.3942	.2128	.1686	.2667
8	.80	.3715	.1750	.1403	.3504
9	.90	.4603	.1662	.1476	.2874

https://doi.org/10.1017/sjp.2015.24 Published online by Cambridge University Press

assumption is explicitly violated; if the comparison favors a BPM, then the conclusion is that invariance is violated. The latter, inspired by the IRT literature, consists in partitioning the observed data set into two groups (one containing all patterns below a certain cutoff, and one containing all patterns above the cutoff), to fit the model in each of them, and to apply some statistical test to evaluate the difference between the parameter estimates of the two groups. If the test is statistically significant, then the parameter invariance is violated. This second method, called restricted naïve test, does not work properly, because it leads to biased parameter estimates in both groups. Indeed this bias is a direct effect of the manipulations introduced to partition the data into the two groups, and says nothing about possible departures of the data from the parameter invariance assumption.

In the present work, the analysis was extended to a more general method for constructing the two groups. The groups are formed by choosing a proportion $p^{\uparrow}$ of the patterns above a certain cutoff c, and a proportion $p^{\downarrow}$ of the patterns below c. Theoretical results, simulations and an empirical application showed that, also the general method suffers of the same problems as the restricted naïve test. Again, the manipulations of the data that one implements for setting up the two groups, lead to biased parameter estimates that have nothing to do with the violation of the invariance assumption. Given these observations, the only available method for testing the BLIM's invariance is, currently, bipartition models.

References

Andersen E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, *38*, 123–140. http://dx.doi.org/10.1007/BF02291180

de la Torre J., & Lee Y.-S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement*, *47*, 115–127. http://dx.doi.org/10.1111/j.1745-3984.2009.00102.x

de Chiusole D., Anselmi P., Stefanutti L., & Robusto E. (2013). The gain-loss model: Bias and variance of the parameter estimates. *Electronic Notes in Discrete Mathematics*, *42*, 33–40.

de Chiusole D., & Stefanutti L. (2013). Modeling skill dependence in probabilistic competence structures. *Electronic Notes in Discrete Mathematics*, *42*, 41–48. http://dx.doi.org/10.1016/j.endm.2013.05.144

de Chiusole D., Stefanutti L., Anselmi P., & Robusto E. (2013). Assessing parameter invariance in the BLIM: Bipartition Models. *Psychometrika*, *78*, 710–724. http://dx.doi.org/10.1007/s11336-013-9325-5

Doignon J.-P., & Falmagne J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, *23*, 175–196. http://dx.doi.org/10.1016/S0020-7373(85)80031-6

Doignon J.-P., & Falmagne J.-C. (1999). *Knowledge Spaces*. Berlin, Heidelberg, and New York, NY: Springer-Verlag.

Falmagne J.-C., & Doignon J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, *41*, 1–23. http://dx.doi.org/10.1111/j.2044-8317.1988.tb00884.x

Falmagne J.-C., & Doignon J.-P. (1988b). A Markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology*, *32*, 232–258. http://dx.doi.org/10.1016/0022-2496(88)90011-9

Falmagne J.-C., & Doignon J.-P. (2011). *Learning spaces*. New York, NY: Springer.

Glas C. A. W., & Verhelst N. D. (1995). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar, editors, *Rasch Models: Foundations, Recent Developments, and Applications*. New York, NY: Springer.

Heller J., Stefanutti L., Anselmi P., & Robusto E. (2014). Cognitive diagnostic models and knowledge space theory. The non-missing link. *Manuscript submitted for publication*.

Heller J., & Wickelmaier F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, *42*, 49–56. http://dx.doi.org/10.1016/j.endm.2013.05.145

Junker B. W., & Sijtsma K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272. http://dx.doi.org/10.1177/01466210122032064

Robusto E., Stefanutti L., & Anselmi P. (2010). The Gain-Loss Model: A probabilistic skill multimap model for assessing learning processes. *Journal of Educational Measurement*, *47*, 373–394. http://dx.doi.org/10.1111/j.1745-3984.2010.00119.x

Schrepp M. (2005). About the connection between knowledge structures and latent class models. *Methodology*, *1*, 93–103. http://dx.doi.org/10.1027/1614-2241.1.3.93

Spoto A., Stefanutti L., & Vidotto G. (2012). On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures. *Journal of Mathematical Psychology*, *56*, 248–255. http://dx.doi.org/10.1016/j.jmp.2012.05.001

Stefanutti L., Heller J., Anselmi P., & Robusto E. (2012). Assessing local identifiability of probabilistic knowledge structures. *Behavior Research Methods*, *44*, 1197–1211. http://dx.doi.org/10.3758/s13428-012-0187-z

Stefanutti L., & Robusto E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74*, 83–96. http://dx.doi.org/10.1007/s11336-008-9095-7