Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Psicologia Applicata

SCUOLA DI DOTTORATO DI RICERCA IN: Scienze Psicologiche

INDIRIZZO: Psicologia Sociale e della Personalità

CICLO: XXIII

The Gain-Loss Model: A formal model for assessing learning processes

Direttore della Scuola: Ch.mo Prof. Clara Casco

Supervisore: Ch.mo Prof. Egidio Robusto

Dottorando: Pasquale Anselmi

Allo zio Pino

Contents

Introduction

Traditionally, educational assessment has evaluated the knowledge of students at the end of a course or a program by assigning a score that summarizes their learning. In the last few decades, however, advances in the fields of human learning and performance have strongly encouraged the development of assessment practices that focus on the specific knowledge and skills required by students to perform a task. A new approach to assessment has therefore started to emerge which is called *formative*, because it evaluates the specific skills of students in order to guide teaching and learning. At the same time, it provides information concerning the effectiveness of educational interventions in promoting specific learning.

The thesis presents the Gain-Loss Model (GaLoM), which is a formal model for assessing learning processes. The theoretical framework is knowledge space theory, a novel approach to the assessment of knowledge proposed by Doignon and Falmagne in 1985. In contrast with the traditional approach, which is based on the numerical evaluation of some “aptitude”, knowledge space theory provides a non-numerical, but nevertheless precise representation of the knowledge of students in a certain domain. Such a representation is consistent with the aims of formative assessment.

The GaLoM assesses the knowledge of students in the different steps of the learning process, and the effectiveness of an educational intervention, referred to as a learning object, in promoting specific learning. The core element is represented by a skill multimap associating each problem with a collection of subsets of skills that are necessary and sufficient to solve it. Model parameters are initial probabilities of the skills, effects of learning objects on gaining and losing the skills, careless error and lucky guess probabilities of the problems.

The model has been the subject of investigation at different levels. Its functioning has been analyzed under different conditions, and theoretical developments have been proposed for improving its informative power in practical applications. The investigations have been conducted through simulated studies and empirical applications. Both kinds of studies have been used because they are very informative when a new model is developed. Simulated studies allow the model to be tested in situations in which all information concerning the data which is relevant to the analysis (e.g., the association between problems and skills underlying the data, the noise of the data, and the value of the true parameters) is present. Empirical applications allow the model to be tested when dealing with all the elements of uncertainty that characterize the use of a formal model in practice.

The thesis presents the work completed on the model. On one hand, the theoretical development of the model itself, as well as some extensions of it, are described. On the other hand, the results of the simulation studies and the empirical applications are presented. The argumentation develops in the following way.

Chapter 1 provides a brief introduction to knowledge space theory. The general idea and the basic elements are presented. Emphasis is placed on the concepts which represent the theoretical background of the GaLoM.

Chapter 2 presents the GaLoM. The mathematical specification of the model and the estimation of its parameters are described. The results are presented of a simulation study which investigated the characteristics of identifiability and goodness-of-recovery of model parameters under different conditions. Moreover, the results of an empirical application are described, which tested the capability of the model to distinguish between different educational interventions.

Chapter 3 presents an extension of the GaLoM in which the estimates of some model parameters are constrained. Results are described of a simulation study which investigated the usefulness of the proposed extension for reducing the variability of the whole set of parameters in non-identifiable models, and for facilitating the identification of the skill multimap which underlies the data.

Chapter 4 tests the usefulness of the aforementioned extension of the model for identifying, from a number of alternative solutions, the skill multimap that best approximates the one underlying the data. The results of a simulation study and an empirical application are presented.

Chapter 5 presents a logistic reparametrization of the GaLoM. It shows how logistic parameters can be computed, which provide a new way of reading information concerning the learning process. The usefulness of the reparametrization for highlighting features of the skills and of the learning objects is illustrated through an empirical application.

Chapter 6 concludes the argumentation. The most important results are reviewed, together with some suggestions for increasing the usefulness of the model which derive from them. Some suggestions for further development of the model are also presented.

Chapter 1

Knowledge structures for the assessment of knowledge

Much of educational assessment aims to evaluate the learning of a student after the teaching is over. This assessment is known as *summative*, because it assigns each student a score that summarizes his learning outcome at the end of a course or a program. Well-known tests, such as the Scholastic Assessment Test and the Graduate Record Examination are currently administered in many countries to inform decisions concerning admissions, placements, scholarships and fellowships. The analyses are often performed by means of models proposed within classical testing theory and item response theory because they are useful for grading the students on continuous unidimensional scales. The success of these models has been strongly favored by the availability of the computing machines required for handling the analyses.

In the last few decades, advances in the fields of human learning and performance have highlighted the cognitively complex nature of domains such as mathematics, science and writing (Nichols & Sugrue, 1999), and they have strongly encouraged the development of assessment practices that focus on "component skills and discrete bits of knowledge to encompass the more complex aspects of student achievement" (Pellegrino, Chudowsky, & Glaser, 2001, p. 3). A new approach to assessment has therefore started to emerge which is called *formative* because it evaluates the specific knowledge, skills and abilities of students in order to guide teaching and learning (DiBello & Stout, 2007). Formative assessment enables teachers to determine if students have learned what they were supposed to, as well as ascertaining whether the educational intervention has been effective in promoting specific learning or not. Depending on the results of the assessment, further steps of teaching and learning are planned. Models for formative assessment pinpoint students' specific strengths and weaknesses by assigning multidimensional skill profiles to them. No numerical representation of students is obtained. The development of these models has been made possible by the availability of much more powerful computing machines.

This thesis presents a formal model for assessing the knowledge of students and the effectiveness of educational interventions. The theoretical framework is knowledge space theory, which is a novel approach for the assessment of knowledge proposed by Doignon and Falmagne in 1985.

Knowledge space theory represents a sharp departure from the traditional approach, which is based on the numerical evaluation of some "aptitude". This approach is based on the nineteenth century position that physics, and the associated methodological approach, was the model for other sciences to imitate. Such a position is still influential today with a detrimental effect on fields, such as the psychological sciences, whose phenomena of interest are of a different nature than those of physics. Falmagne, Cosyn, Doignon, and Thiéry (2006) provided a nice illustration of the limitations of a purely numerical description of some phenomena by an analogy with sports. The full quotation is:

> "It is true that the success of an athlete in a particular sport is often described by a set of impressive numbers. So, imagine that some committee of experts has carefully designed an 'Athletic Quotient'

> or 'A.Q.' test, intended to measure athletic prowess. Suppose that three exceptional athletes have taken the test, say Michael Jordan, Tiger Woods and Pete Sampras. Conceivably, all three of them would get outstanding A.Q.'s. But these high scores equating them would completely misrepresent how essentially different from each other they are. One may be tempted to salvage the numerical representation and argue that the assessment, in this case, should be multidimensional. However, adding a few numerical dimensions capable of differentiating Jordan, Woods and Sampras would only be the first step in a sequence. Including Greg Louganis or Pele to the evaluated lot would require more dimensions, and there is no satisfactory end in sight. Besides, assuming that one would settle for a representation in n dimensions, for some small n equal 3, 4 or 5 say, the numerical vectors representing these athletes would be poor, misleading expressions of the exquisite combination of skills making each of them a champion in his own specialty." (Falmagne et al., 2006, p. 63).

Knowledge space theory is different in spirit from the traditional approach to the assessment of knowledge and, consistently with the aims of formative assessment, it provides a non-numerical, multidimensional representation of the characteristics of a student.

Since it was first formulated, knowledge space theory has accumulated an impressive body of theoretical results and developments. The most important of these are reviewed in Doignon and Falmagne (1999), and more recently in Falmagne and Doignon (2011). Moreover, it has been applied in a number of different domains, such as elementary stochastic calculus, geometry, continuation of number and letter series, chess and, surprisingly, sport. These applications are collected in a book edited by (Albert & Lukas, 1999). However, the most significant application is probably the one concerning the computer system ALEKS (acronym for Assessment and LEarning in Knowledge Spaces). ALEKS is an internet-based, automated mathematics tutor which adaptively assesses students' knowledge in subjects from basic mathematics to precalculus. It is currently used for supporting teaching and learning in many countries (for details, see Falmagne & Doignon, 2011; Falmagne et al., 2006).

This chapter is a brief introduction to knowledge space theory. The general idea and the basic elements are outlined. In particular, the concepts are introduced which represent the background of the model that is presented in this thesis. No mathematical formalization is given here. First, the two basic concepts of knowledge state and knowledge structure are introduced. Then, different methods for building a knowledge structure in a given domain are reviewed. Emphasis is placed on a method which takes into account the skills underlying the responses to the problems. Finally, there is a description of how a knowledge structure can be empirically validated in a probabilistic framework. The reader interested in further exploring the theoretical concepts is referred to Falmagne and Doignon (2011) and to the specific literature which will be indicated when appropriate.

1.1 Knowledge states and knowledge structures

Knowledge space theory is based on the notion of a field of knowledge that is parsed into a possibly large set of "units of knowledge". One such unit might be a specific problem. In the field of high school algebra, for instance, a problem concerning quadratic equations, such as $3x^2+5x-2=0$. A knowledge domain is therefore identified with a set Q of problems, each of which has a correct response.

A first key concept is that of knowledge state, which is the collection $K \subseteq Q$ of all the problems in the domain Q that a student is capable of solving. Not all the $2^{|Q|}$ subsets of Q are plausible knowledge states. In general, the solution behaviors in a knowledge domain will exhibit some dependencies. From the responses to some problems, inferences can sometimes be made regarding the responses to other problems. These mutual dependencies in the solution behaviour determine which subsets of Q are feasible knowledge states.

A second key concept is that of knowledge structure, which is a collection $\mathcal{K}$ of knowledge states K of a knowledge domain Q. A knowledge structure is assumed to contain at least the empty set $\emptyset$ and the full set Q, as it may always be the case that none or all the problems are solved. There are different kinds of knowledge structures. A structure that is closed under union (i.e., the union of any two states is also a state of the structure) is a knowledge space. A structure that is closed under intersection (i.e., the intersection of any two states is also a state of the structure) is a simple closure space. A structure that is closed under both union and intersection is a quasi ordinal space.

The aforementioned concepts are illustrated by an example. Consider a domain Q that contains five problems and is equipped with the knowledge structure

$$\mathcal{K} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{2,3\}, \{1,2,4\}, \{1,3,5\}, Q\}$$

The knowledge structure $\mathcal{K}$ can be graphically represented by ordering its knowledge states by set inclusion as in Figure 1.1. This representation is known as a Hasse diagram. In this example, not all subsets of Q have been assumed to be knowledge states. In fact, the knowledge structure $\mathcal{K}$ contains eight states out of thirty-two possible ones. Moreover, the structure is a simple closure space because it is closed under intersection but not under union (e.g., $\{1\} \cup \{2\}$ is not a state of $\mathcal{K}$).

1.2 How to build a knowledge structure

In applying knowledge space theory to a field of knowledge, the most important stage lies in building the structure, that is, in finding out what the feasible

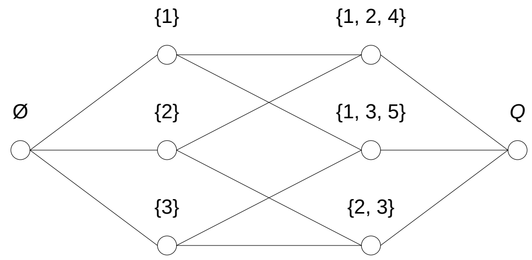

Figure 1.1: Hasse diagram of the knowledge structure $\mathcal{K} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{2,3\}, \{1,2,4\}, \{1,3,5\}, Q\}$.

knowledge states are. The knowledge structure, indeed, represents the core element which determines the quality of the assessment of knowledge.

Different methods have been suggested for building a knowledge structure. They can be grouped into three classes according to the fact that the knowledge structure is:

1. Extracted from the data collected on a sample of persons;

2. Derived from questioning experts in the knowledge domain under study;

3. Derived from considering the cognitive processes underlying the responses to the problems.

There are essentially two methods which extract the structure from the data. A first method is based on counting the frequencies of all the response patterns, and on taking as feasible knowledge states the response patterns whose observed frequencies exceed a specific threshold. This method has been illustrated, for example, by Falmagne (1989), and Schrepp (1999a). A limit of this method is that it is feasible only when enough response patterns are available. A slightly different method is based on item tree analysis (Van Leeuwe, 1974), in which logical implications between the problems are derived from an inspection of empirical contingency tables for all possible pairs of problems. Developments and applications of item tree analysis have been described, for example, by Held and

Korossy (1998), Sargin and Ünlü (2009), and Schrepp (1999b, 2002, 2003). An important limit of this method is that the knowledge structure resulting from the analysis strongly depends on the choice of appropriate parameters values (e.g., frequency thresholds).

The methods which are based on the questioning of experts represent the typical way in which a knowledge structure is built. The experts are not directly asked about what the knowledge states are because the knowledge states are something that is abstract and difficult to convey in an exact manner. Rather, experts' awareness of the feasible knowledge states is indirectly derived by asking questions such as: "Suppose that a student is not capable of solving problem q. Could he nevertheless solve problem q'?", and: "Suppose that a student is not capable of solving problems $q_1, q_2, \ldots, q_n$. Could he nevertheless solve problem q'?". From the responses to questions of the first type, a quasi ordinal space is derived, whereas from the responses to questions of the second type, a knowledge space is derived. The algorithms for computerized querying procedures have been described by Falmagne, Koppen, Villano, Doignon, and Johannesen (1990), Koppen and Doignon (1990), and Messick (1989). Refinements and applications have been reported by Cosyn and Thiéry (2000), Dowling (1993), Kambouri, Koppen, Villano, and Falmagne (1994), and Koppen (1993). This method for determining a knowledge structure is tedious and time consuming, even for domains of moderate size. Moreover, it is highly sensible to experts' mistakes and does not guarantee the validity of the knowledge structure that is obtained. Response data collected on a sample of students have to be used to test and refine the knowledge structure.

There are essentially two methods which derive a knowledge structure from explicit hypotheses and assumptions about the cognitive processes underlying the responses to the problems. A first method is based on task analysis, that is the systematic examination of the content of the problems for extracting the relevant components. Each component is associated with a demand of the problem, and it is supposed to contribute to its difficulty. The knowledge structure is then derived from explicitly formulated principles about the component combinations contained in each problem. Details about this method, together with a number of applications, can be found in Albert and Held (1994), Albert and Lukas (1999), and Lukas and Albert (1993). A second method is based on explicit assumptions about the skills which underlie the responses to the problems. Each problem is associated with the skills that are assumed to be useful or instrumental to solve it, and the feasible knowledge states are derived from this association. Applications of this method can be found in Albert and Lukas (1999), Düntsch and Gediga (1995), and Gediga and Düntsch (2002). The structures derived by task analysis or skill assignment have to be tested on response data.

An important feature of the last two methods is that they take into account theoretical terms that may be interpreted psychologically, and formulates explicitly the link between demands or skills and the feasible knowledge states. The knowledge structures derived from psychological theories about the cognitive processes involved in the response process may be suitable to test the theories empirically.

1.3 Skills and knowledge structures

In its original formalization, knowledge space theory was purely behavioristic. Its focus was explicitly on the observable solution behaviour at the level of the problems.

Following Falmagne et al. (1990), however, the scope of the theory was extended by taking into account the cognitive processes which underlie the observable solution behaviour. In subsequent independent developments, Doignon (1994), Düntsch and Gediga (1995), Heller and Repitsch (2008), and Korossy (1999) proposed various approaches to consider the skills and competencies that are required to perform a task. Several results obtained in this area can be found in Albert and Lukas (1999).

The approach described here is based on the work by Doignon (1994), and Falmagne et al. (1990). The assumption is made that there are some basic skills which may consist in methods, notions, abilities, solution procedures, and so on. Each problem is associated with the skills that are useful or instrumental to solve it, and the feasible knowledge states are derived from this association. There are three methods to formalize the linkage between skills and problems: the conjunctive model, the disjunctive model, and the competency model.

In the conjunctive model, all the skills associated with a problem are required in order to solve it. In the disjunctive model, only one of the skills associated with a problem suffices to solve it. The knowledge structures delineated via the conjunctive model are simple closure spaces, whereas those delineated via the disjunctive model are knowledge spaces. The knowledge structures delineated via the conjunctive and the disjunctive model are dual to each other.

As an example, consider four problems and three skills. Consider also that each problem is associated with the skills that are relevant for solving it in accordance with the assignment represented in Table 1.1.

Table 1.1: Skill Assignment in the Conjunctive, Disjunctive and Competency Model

Conjunctive and disjunctive model		Competency model	
Problem	Skills	Problem	Competencies
1	$\{a, b\}$	1	$\{a, b\}, \{a, c\}$
2	$\{c\}$	2	$\{c\}$
3	$\{a, b, c\}$	3	$\{a\}, \{b, c\}$
4	$\{b\}$	4	$\{b\}$

Note. Letters from a to c refer to the three skills.

In the example, problems 2 and 4 are associated with one skill, problem 1 is associated with two skills, and problem 3 is associated with three skills. With three skills there are $2^3 = 8$ different subsets of skills. Each subset of skills represents the hypothetical skill profile of a student. Table 1.2 lists the knowledge states delineated by each subset of skills via the conjunctive and disjunctive models. Note that the empty subset of skills always delineates the empty knowledge state $\emptyset$, and that the full subset of skills always delineates the full knowledge state Q.

Table 1.2: Knowledge States Delineated by each Subset of Skills via the Conjunctive, Disjunctive and Competency Model

Subset of skills	K_c	K_d	K_{co}
$\{\}$	$\emptyset$	$\emptyset$	$\emptyset$
$\{a\}$	$\emptyset$	$\{1,3\}$	$\{3\}$
$\{b\}$	$\{4\}$	$\{1,3,4\}$	$\{4\}$
$\{c\}$	$\{2\}$	$\{2,3\}$	$\{2\}$
$\{a,b\}$	$\{1,4\}$	$\{1,3,4\}$	$\{1,3,4\}$
$\{a,c\}$	$\{2\}$	$\{1,2,3\}$	$\{1,2,3\}$
$\{b,c\}$	$\{2,4\}$	Q	$\{2,3,4\}$
$\{a,b,c\}$	Q	Q	Q

Note. Letters from a to c refer to the three skills. K_c, K_d and K_{co} = knowledge state delineated via the conjunctive, disjunctive and competency model, respectively.

The knowledge structure delineated via the conjunctive model is $\mathcal{K}_c = \{\emptyset, \{2\}, \{4\}, \{1,4\}, \{2,4\}, Q\}$, the one delineated via the disjunctive model is $\mathcal{K}_d = \{\emptyset, \{1,3\}, \{2,3\}, \{1,2,3\}, \{1,3,4\}, Q\}$. Note that $\mathcal{K}_c$ is a simple closure space, that $\mathcal{K}_d$ is a knowledge space, and that $\mathcal{K}_c$ and $\mathcal{K}_d$ are dual to each other.

The conjunctive model and the disjunctive model are particular cases of a more general model. In the competency model, each problem is associated with a collection of subsets of skills. Each subset of skills is called competency, and it represents a method or a strategy for solving the problem. Possessing just one of these competencies is sufficient for solving the problem, but all the skills contained in it are necessary. The knowledge structure delineated by a skill multimap is a general structure, that is, it is not necessarily closed under union or intersection.

As an example, consider four problems and three skills. Each problem is associated with its competencies according to Table 1.1. Problems 2 and 4 are associated with one competency, and problems 1 and 3 are associated with two competencies. The knowledge states delineated by each subset of skills via the competency model are listed in Table 1.2. The knowledge structure is $\mathcal{K}_{co} = \{\emptyset, \{2\}, \{3\}, \{4\}, \{1,2,3\}, \{1,3,4\}, \{2,3,4\}, Q\}$.

Note that this structure is closed under neither union (e.g., $\{2\} \cup \{3\}$ is not a state of $\mathcal{K}_{co}$) nor intersection (e.g., $\{1,2,3\} \cap \{1,3,4\}$ is not a state of $\mathcal{K}_{co}$).

1.4 Probabilistic knowledge structures

The knowledge structure is a deterministic model of the organization of knowledge in a certain domain. As such, it does not provide a realistic prediction of the responses that the students give to the problems. A probabilistic framework can be introduced to overcome this limitation.

There are essentially two ways in which probabilities must be considered. Firstly, it is expected that the knowledge states occur with different frequencies within the population of reference. It is therefore reasonable to introduce a probability distribution on the knowledge states. Secondly, it is expected that there is not a perfect correspondence between the knowledge state of a student and his response pattern, that is, between the problems that the student is capable of solving and those he actually solves. A student who is able to solve a problem might be careless in responding, and make an error. In contrast, a student who is not able to solve a problem might guess the correct response. A distinction has therefore to be made between the knowledge state, which is a latent construct, and the response pattern, which is a manifest indicator of the knowledge state. If careless errors and lucky guesses are committed, all the $2^{|Q|}$ subsets of Q might be response patterns generated by the knowledge states in the structure. It is therefore reasonable to introduce conditional probabilities of the response patterns given the knowledge states.

Falmagne and Doignon (1988a) have developed a model of this type, which is known as the Basic Local Independence Model (BLIM). The model specifies the probabilities of the response patterns as a function of the probabilities of the knowledge states and of the conditional probabilities of the response patterns, given the knowledge states. The conditional probabilities are governed by the careless error and lucky guess probabilities of the problems on the basis of the assumption that the responses to the problems are locally independent, given the knowledge state of the student. The goodness-of-fit of the model to the data can be assessed by standard statistics, such as Pearson's Chi-square and the likelihood ratio.

The BLIM is the fundamental model in knowledge space theory. Applications have been described, for example, by Falmagne et al. (1990), Stefanutti (2006), and Taagepera, Potter, Miller, and Lakshminarayan (1997). The model has been the foundation of several approaches to the assessment of knowledge and learning. Some of them are reviewed in Doignon and Falmagne (1999), and Falmagne and Doignon (2011). The model presented in this thesis is based on the BLIM as well.

Chapter 2

Assessing learning processes with the Gain-Loss Model

Formative assessment evaluates specific knowledge of students and the effectiveness of the educational interventions in promoting specific learning. Depending on the results of the assessment, further steps of teaching and learning are planned.

Within the context of formative assessment, the Gain-Loss Model (GaLoM, Robusto, Stefanutti, & Anselmi, 2010; Stefanutti, Anselmi, & Robusto, in press) is presented, which is a formal model for assessing learning processes. The GaLoM assesses the knowledge of students in the different steps of the learning process, and the effectiveness of the educational intervention in promoting specific learning. Such assessments can be obtained with respect to the classroom and to individual students. Therefore, the model provides the teacher with diagnostic information on two different levels. At the classroom level, it informs the teacher about the effectiveness of the educational program that has been carried out and of the testing tool (e.g., a collection of problems) that has been used to assess the knowledge. Moreover, at the student level, the model provides information that enables the teacher to select the best educational intervention for the specific weaknesses of each student.

The chapter is organized as follows. In the next paragraph, the GaLoM is introduced, and the aspects concerning mathematical specification and parameter estimation are described. Next, a simulation study that assesses identifiability and goodness-of-recovery of the model is provided. Then, an empirical application that tests the capability of the model of distinguishing between an effective and an ineffective educational intervention is presented. Finally, practical benefits for teaching and learning that derive from using the model are explored and discussed, and a comparison with other models presented in literature is provided.

2.1 The Gain-Loss Model

In knowledge space theory (Doignon & Falmagne, 1985, 1999), the knowledge of a student in a particular domain is operazionalized as the observable solution behaviour of that student on a specific set of problems. In contrast, in the present approach the knowledge of a student is conceptualized in terms of the unobservable skill profile which characterizes that student and which accounts for his observable solution behaviour. According to the competence-performance conception (Korossy, 1999), the two ways of conceptualizing knowledge respectively concern levels of performance and competence. Given a collection Q of problems, the performance state of a student is the collection $K \subseteq Q$ of all the problems that this student is capable of solving. A performance structure is a pair $(Q, \mathcal{K})$, where $\mathcal{K}$ is a collection of subsets of Q. Given a collection S of skills, the competence state of a student is the collection $C \subseteq S$ of non directly observable skills possessed by him and which underlie his observable responses to the problems. A competence structure is then a pair $(S, \mathcal{C})$, where $\mathcal{C}$ is a collection of subsets of S. In the sequel, the performance structure and the competence structure will be just indicated with $\mathcal{K}$ and $\mathcal{C}$[1].

[1]The performance state and the performance structure correspond to what Doignon and Falmagne denote with knowledge state and knowledge structure.

The GaLoM focuses on the skills that the students must possess in order to solve the problems. The skills are elementary and indecomposable units of knowledge that may represent either declarative or procedural knowledge, including notions, abilities, solution procedures, tricks, and so on.

The GaLoM models the learning process of the students as a function of the interaction between their competence state and the effect of an educational intervention, called learning object. The term *learning object* is used here to represent every didactic intervention that supports learning and has the potential capability of changing the competence state of the students. Learning objects could be general courses on particular topics, instructional content (including texts, web pages, images, sounds, videos), glossaries of terms, quizzes, exercises, case studies, educational games, and so on. The model assesses the effect of learning objects on the attainment of specific skills which are required to solve problems in a given field of knowledge. The relation between skills and problems is established by means of a competency model (Doignon, 1994; Doignon & Falmagne, 1999), in which a skill multimap associates each problem with a collection of subsets of skills that are necessary and sufficient to solve it.

The GaLoM is characterized by five types of parameters. The parameter concerning the initial probability of the skills specifies what skills the students possess before the teaching begins. The gain and loss parameters respectively specify if the students attain and eventually lose specific skills as a result of the learning object they have been presented with. The careless error parameter specifies whether the students who master a problem fail it through inattention, whereas the lucky guess parameter specifies whether the students who do not master a problem solve it by guessing.

2.1.1 Mathematical specification of the Gain-Loss Model

The GaLoM assesses the effect of m different learning objects in processes with two assessement steps. The experimental design is $t_1 \rightarrow I_o \rightarrow t_2$, where I represents the educational intervention carried out with learning object $o \in \{1, 2, \ldots, m\}$, and t_1 and t_2 are the pretest and posttest, respectively.

Let S be a finite and non-empty set of discrete skills, and C be any subset of S. C represents the unknown competence state of a student. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two discrete random variables whose realizations are the competence states of a student at the pretest and posttest, respectively. Let Q be a finite and non-empty set containing n dichotomous problems, and $\mathbf{R}_1$ and $\mathbf{R}_2$ be two discrete random variables whose realizations are the response patterns $\mathbf{r} \in \{0, 1\}^n$ of a student at the pretest and posttest, respectively.

Assumptions of the model are:

1. The response patterns $\mathbf{R}_1$ and $\mathbf{R}_2$ are locally independent, given the states $\mathbf{C}_1$ and $\mathbf{C}_2$: $P(\mathbf{R}_1, \mathbf{R}_2|\mathbf{C}_1, \mathbf{C}_2) = P(\mathbf{R}_1|\mathbf{C}_1)P(\mathbf{R}_2|\mathbf{C}_2)$;

2. The initial state $\mathbf{C}_1$ does not depend on learning object o;

3. State $\mathbf{C}_2$ depends on previous state $\mathbf{C}_1$ and on learning object o.

It follows that the conditional probability that $\mathbf{r}_1$ and $\mathbf{r}_2$ are the response patterns of a randomly sampled student at the pretest and posttest, given learning object o, is:

$$P(\mathbf{R}_1 = \mathbf{r}_1, \mathbf{R}_2 = \mathbf{r}_2|o) = \sum_{C \subseteq S} \sum_{D \subseteq S} P(\mathbf{R}_1 = \mathbf{r}_1|\mathbf{C}_1 = C)P(\mathbf{R}_2 = \mathbf{r}_2|\mathbf{C}_2 = D)$$
$$P(\mathbf{C}_2 = D|\mathbf{C}_1 = C, o)P(\mathbf{C}_1 = C), \quad (2.1)$$

where $P(\mathbf{C}_1 = C)$ is the initial probability of the state C at the pretest, $P(\mathbf{C}_2 = D|\mathbf{C}_1 = D|\mathbf{C}_1 = C, o)$ is the transition probability from state C at the pretest to state D at the posttest, $P(\mathbf{R}_1 = \mathbf{r}_1|\mathbf{C}_1 = C)$ and $P(\mathbf{R}_2 = \mathbf{r}_2|\mathbf{C}_2 = D)$ are the emission probabilities of response patterns $\mathbf{r}_1$ and $\mathbf{r}_2$ at the pretest and posttest, respectively.

Equation 2.1 is the basic equation of the model. Assuming total independence among the skills in S, the probability $P(\mathbf{C}_1 = C)$ is resolved according to Equation 2.2:

$$P(\mathbf{C}_1 = C) = \prod_{s \in S} \pi_s^{w(s,C)} (1 - \pi_s)^{1-w(s,C)}, \tag{2.2}$$

where $\pi_s = P(s \in \mathbf{C}_1)$ is the probability that skill s belongs to the initial competence state, and $w(s, C) \in \{0, 1\}$ is equal to 1 if skill s belongs to state C.

Given the independence among the skills modelled at time t_1, the presence or absence of a skill at time t_2 only depends on its presence or absence at a previous time. Note that the assumption of independence among the skills is not a necessary condition for the models to be applied.

Let gain γ_{os} be the probability $P(s \in \mathbf{C}_2 | s \notin \mathbf{C}_1, o)$ that the students presented with learning object o gain the skill s going from the pretest to the posttest, and let loss λos be the probability $P(s \notin \mathbf{C}_2 | s \in \mathbf{C}_1, o)$ that the same students lose it. The conditional probability of state D at the posttest, given state C at the pretest and learning object o, turns out to be:

$$P(\mathbf{C}_2 = D | \mathbf{C}_1 = C, o) = \prod_{s \in S} \left[\lambda_{os}^{1-w(s,D)} (1 - \lambda_{os})^{w(s,D)} \right]^{w(s,C)}$$
$$\left[\gamma_{os}^{w(s,D)} (1 - \gamma_{os})^{1-w(s,D)} \right]^{1-w(s,C)},$$

where $w(s, C) \in \{0, 1\}$ (resp. $w(s, D)$) is equal to 1 if skill s belongs to state C (resp. D).

Via the competency model, a skill multimap associates each problem with a collection of subsets of skills that are necessary and sufficient to solve it (Doignon, 1994; Doignon & Falmagne, 1999). Each subset of skills is called *competency*. A skill multimap is a triple (Q, S, δ), where δ is a mapping from Q to the powerset of 2^S such that $\delta(q) \neq \emptyset$ for each $q \in Q$ (i.e., each problem is associated with at least one competency), and $C \neq \emptyset$ for each $C \in \delta(q)$ (i.e., each competency contains at least one skill). Each subset C in $\delta(q)$ represents a competency for q. The performance state delineated by $C \subseteq S$ via the competency model is specified by:

$$M(C) = \{q \in Q : C' \subseteq C \text{ for some } C' \in \delta(q)\},$$

and the performance structure delineated by the skill multimap (Q, S, δ) is:

$$\mathcal{K} = \{M(C) | C \subseteq S\}.$$

The performance structure delineated by a skill multimap is a general structure, that is, it is not necessarily closed under union or intersection (Doignon & Falmagne, 1999).

Let careless error α_q be the probability $P[\mathbf{R}_q = 0 | q \in M(\mathbf{C})]$ that the students fail problem q given that they possess a competence for it, and let lucky guess β_q be the probability $P[\mathbf{R}_q = 1 | q \notin M(\mathbf{C})]$ that the same students solve problem q given that they do not possess any competence for it. Therefore, careless error and lucky guess are the error parameters that govern the emission probabilities of response patterns. Assuming responses to the problems are locally independent[2], given the student's competence state, the conditional probability of response pattern $\mathbf{r}$, given state C, is:

$$P(\mathbf{R}_t = \mathbf{r} | \mathbf{C}_t = C) = \prod_{q=1}^{n} \left[\alpha_q^{1-r_q}(1-\alpha_q)^{r_q}\right]^{v(q,C)} \left[\beta_q^{r_q}(1-\beta_q)^{1-r_q}\right]^{1-v(q,C)},$$

where $t \in \{1, 2\}$, $v(q, C) \in \{0, 1\}$ is equal to 1 if problem q is solvable by state C, and $r_q \in \{0, 1\}$ is equal to 1 if problem q is solved.

[2]Local independence specifies that, conditional on the competence states of the student, the responses to the problems are independent. That is, any correlation between responses is completely explained by the states. Local independence is the basic assumption underlying latent variable models, such as factor analysis, latent trait analysis, and latent class analysis. Details can be found in Lazarsfeld and Henry (1968).

2.1.2 Parameter estimation of the Gain-Loss Model

In the GaLoM, an initial probability π_s is estimated for each skill $s \in S$, a gain parameter γ_{os} and a loss parameter λ_{os} for each learning object $o \in \{1, 2, \ldots, m\}$ and each skill s, a careless error parameter α_q and a lucky guess parameter β_q for each problem $q \in Q$. Therefore, the total number of parameters to be estimated is $2n + (1 + 2m)|S|$, where n is the number of problems, m is the number of learning objects, and $|S|$ is the cardinality of the set of skills. The MATLAB code for estimating and testing the GaLoM is provided in Appendix A.1.

Since the GaLoM is essentially a latent class model where the latent classes are the competence states in $\mathcal{C}$, maximum likelihood estimates of the parameters (Stefanutti et al., in press) can be computed by an application of the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977).

Let $\mathbf{X}_t$ be a $i \times n$ binary matrix each row of which is the response pattern of student $j \in \{1, 2, \ldots, i\}$ to the n problems in Q at time $t \in \{1, 2\}$. Each cell $x_{tjq} \in \{0, 1\}$ of $\mathbf{X}_t$ is equal to 1 if student j solved problem $q \in \{1, 2, \ldots, n\}$ at time t. The observed data sample is the binary matrix $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ with dimensions $i \times 2n$. Let $\boldsymbol{\theta} = (\alpha, \beta, \pi, \gamma, \lambda)$ be the vector of all model parameters, and let $\mathbf{o} = (o_1, o_2, \ldots, o_i)'$ be the vector that associates each student to the learning object he was presented with. The observed likelihood of $\mathbf{X}_1$ and $\mathbf{X}_2$, given the model parameters $\boldsymbol{\theta}$ and the learning objects specified by $\mathbf{o}$, is derived from 2.1. Assuming multinomial sampling, it takes on the form:

$$l(\mathbf{X}_1, \mathbf{X}_2|\boldsymbol{\theta}, \mathbf{o}) = \prod_{j=1}^{i} \left[\sum_{C \subseteq S} \sum_{D \subseteq S} P(\mathbf{x}_{1j}|C) P(\mathbf{x}_{2j}|D) P(D|C, o_j) P(C) \right],$$

where $\mathbf{x}_{1j}$ (resp. $\mathbf{x}_{2j}$) is the jth row-vector of matrix $\mathbf{X}_1$ (resp. $\mathbf{X}_2$).

The competence states of the students are obviously unknown, but, if there was complete information, each student j would be represented by a quadruple $(\mathbf{x}_{1j}, \mathbf{x}_{2j}, Y_{1j}, Y_{2j})$, where $\mathbf{x}_{1j}$ (resp. $\mathbf{x}_{2j}$) is a $1 \times n$ binary vector representing the response pattern of j at time 1 (resp. time 2), and Y_{1j} (resp. Y_{2j}) is the competence state of j at time 1 (resp. time 2). Therefore, the complete data sample would be the quadruple $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2)$, and the complete data log-likelihood of the model would turn out to be:

$$\ln l(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2|\boldsymbol{\theta}, \mathbf{o}) = \sum_{j=1}^{i} \ln P(\mathbf{x}_{1j}, \mathbf{x}_{2j}, Y_{1j}, Y_{2j}|\boldsymbol{\theta}, \mathbf{o}),$$

where $P(\mathbf{x}_{1j}, \mathbf{x}_{2j}, Y_{1j}, Y_{2j}|\boldsymbol{\theta}, \mathbf{o})$ is the joint probability of response patterns $\mathbf{x}_{1j}$ and $\mathbf{x}_{2j}$ and of competence states Y_{1j} and Y_{2j} given the model parameters $\boldsymbol{\theta}$ and the learning objects specified by $\mathbf{o}$.

In each iteration of the EM algorithm, the conditional expectation of the complete data log-likelihood is maximized, given the observed data $\mathbf{X}$ and the model parameters $\boldsymbol{\theta}'$ obtained in the previous iteration of the algorithm. Indicating with $U(\boldsymbol{\theta}, \boldsymbol{\theta}')$ this conditional expectation, the following derivation is made:

$$\begin{aligned}
U(\boldsymbol{\theta}, \boldsymbol{\theta}') &= E[\ln l(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2|\boldsymbol{\theta}, \mathbf{o})|\mathbf{X}, \boldsymbol{\theta}'] \\
&= E\left[\sum_{j=1}^{i} \ln P(\mathbf{x}_{1j}, \mathbf{x}_{2j}, Y_{1j}, Y_{2j}|\boldsymbol{\theta}, o_j)|\mathbf{X}, \boldsymbol{\theta}'\right] \\
&= \sum_{j=1}^{i} E[\ln P(\mathbf{x}_{1j}, \mathbf{x}_{2j}, Y_{1j}, Y_{2j}|\boldsymbol{\theta}, o_j)|\mathbf{X}, \boldsymbol{\theta}'] \\
&= \sum_{j=1}^{i} \sum_{C \subseteq S} \sum_{D \subseteq S} [\ln P(\mathbf{x}_{1j}, \mathbf{x}_{2j}, C, D|\boldsymbol{\theta}, o_j)] P(C, D|\mathbf{x}_{1j}, \mathbf{x}_{2j}, o_j, \boldsymbol{\theta}') \\
&= \sum_{j=1}^{i} \sum_{C \subseteq S} \sum_{D \subseteq S} \ln[P(\mathbf{x}_{1j}|C) P(\mathbf{x}_{2j}|D) P(D|C, o_j) P(C)] P(C, D|\mathbf{x}_{1j}, \mathbf{x}_{2j}, o_j, \boldsymbol{\theta}'),
\end{aligned}$$

where o_j is the learning object o that student j has been presented with.

The Bayesian posterior probability of competence state at pretest C and posttest D, given response pattern at pretest $\mathbf{x}_{1j}$ and posttest $\mathbf{x}_{2j}$, learning object o_j, and previous estimates of model parameters $\boldsymbol{\theta}'$, is:

$$\begin{aligned}
P(C, D|\mathbf{x}_{1j}, \mathbf{x}_{2j}, o_j, \boldsymbol{\theta}') &= \frac{P(\mathbf{x}_{1j}, \mathbf{x}_{2j}, C, D|o_j, \boldsymbol{\theta}')}{P(\mathbf{x}_{1j}, \mathbf{x}_{2j}|o_j, \boldsymbol{\theta}')} \\
&= \frac{P(\mathbf{x}_{1j}|C, \boldsymbol{\theta}') P(\mathbf{x}_{2j}|D, \boldsymbol{\theta}') P(D|C, o_j, \boldsymbol{\theta}') P(C|\boldsymbol{\theta}')}{\sum_{C'} \sum_{D'} P(\mathbf{x}_{1j}|C', \boldsymbol{\theta}') P(\mathbf{x}_{2j}|D', \boldsymbol{\theta}') P(D'|C', o_j, \boldsymbol{\theta}') P(C'|\boldsymbol{\theta}')}.
\end{aligned}$$

For the purposes of brevity, in the sequel $P(C, D|\mathbf{x}_{1j}, \mathbf{x}_{2j}, o_j, \boldsymbol{\theta}')$ will be indicated with b_{jCD}.

The conditional expected log-likelihood $U(\boldsymbol{\theta}, \boldsymbol{\theta}')$ can be broken up into three functions in the following way:

$$U(\boldsymbol{\theta}, \boldsymbol{\theta}') = U_1(\boldsymbol{\theta}, \boldsymbol{\theta}') + U_2(\boldsymbol{\theta}, \boldsymbol{\theta}') + U_3(\boldsymbol{\theta}, \boldsymbol{\theta}'), \tag{2.3}$$

where $U_1(\boldsymbol{\theta}, \boldsymbol{\theta}')$, $U_2(\boldsymbol{\theta}, \boldsymbol{\theta}')$, and $U_3(\boldsymbol{\theta}, \boldsymbol{\theta}')$ are defined as:

$$U_1(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{j=1}^{i} \sum_{C \subseteq S} \sum_{D \subseteq S} b_{jCD} \ln P(C),$$

$$U_2(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{j=1}^{i} \sum_{C \subseteq S} \sum_{D \subseteq S} b_{jCD} \ln P(D|C, o_j),$$

and

$$U_3(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{j=1}^{i} \sum_{C \subseteq S} \sum_{D \subseteq S} b_{jCD} [\ln P(\mathbf{x}_{1j}|C) + \ln P(\mathbf{x}_{2j}|D)].$$

Note that U_1 only depends on the initial probabilities π_s of the skills, U_2 only depends on gain γ_{os} and loss λ_{os} parameters, and U_3 only depends on careless error α_q and lucky guess β_q parameters.

Estimation of the initial probabilities of the skills

In each iteration of the EM algorithm, the function $U_1(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is maximized by setting to zero its first partial derivatives with respect to the parameter π_s:

$$\frac{\partial U_1(\boldsymbol{\theta}, \boldsymbol{\theta}')}{\partial \pi_s} = \sum_{j} \sum_{C} b_{jC\cdot} \left[\frac{w(s, C)}{\pi_s} - \frac{1 - w(s, C)}{1 - \pi_s} \right], \tag{2.4}$$

where $b_{jC\cdot} = \sum_D b_{jCD}$, and $w(s, C) \in \{0, 1\}$ is equal to 1 if skill s belongs to state C. By setting the right-hand term of (2.4) to zero, it follows that:

$$(1 - \pi_s) \sum_{j} \sum_{C} b_{jC\cdot} w(s, C) = \pi_s \sum_{j} \sum_{C} b_{jC\cdot} [1 - w(s, C)]. \tag{2.5}$$

By solving (2.5) for π_s, one obtains:

$$\pi_s = \frac{\sum_j \sum_C b_{jC\cdot} w(s, C)}{\sum_j \sum_C b_{jC\cdot}},$$

and, given $\sum_C b_{jC\cdot} = 1$, it follows that:

$$\pi_s = \frac{1}{i} \sum_j \sum_C b_{jC\cdot} w(s, C). \tag{2.6}$$

Equation (2.6) represents the adjustment of the estimates of parameter π_s in each iteration of the EM algorithm.

Estimation of the gain and loss parameters

In each iteration of the EM algorithm, the function $U_2(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is maximized by setting to zero its first partial derivatives with respect to the parameters γ_{os} and λ_{os}. The first partial derivative of $U_2(\boldsymbol{\theta}, \boldsymbol{\theta}')$ with respect to the parameter γ_{os} turns out to be:

$$\frac{\partial U_2(\boldsymbol{\theta}, \boldsymbol{\theta}')}{\partial \gamma_{os}} = \sum_{j \in J_o} \sum_C \sum_D b_{jCD} \left[\frac{[1 - w(s, C)]w(s, D)}{\gamma_{os}} - \frac{[1 - w(s, C)][1 - w(s, D)]}{1 - \gamma_{os}} \right], \tag{2.7}$$

where $J_o = \{j : o_j = o\}$ is the set of the subjects presented with learning object o, and $w(s, C) \in \{0, 1\}$ (resp. $w(s, D)$) is equal to 1 if skill s belongs to state C (resp. D). By setting the right-hand term of (2.7) to zero, it follows that:

$$(1 - \gamma_{os}) \sum_{j \in J_o} \sum_C \sum_D b_{jCD}[1 - w(s, C)]w(s, D) =$$
$$\gamma_{os} \sum_{j \in J_o} \sum_C \sum_D b_{jCD}[1 - w(s, C)][1 - w(s, D)]. \tag{2.8}$$

By solving (2.8) for γ_{os}, one obtains the equation for the adjustment of the estimates of such parameters in each iteration of the EM algorithm:

$$\gamma_{os} = \frac{\sum_C \sum_D b_{oCD}[1 - w(s, C)]w(s, D)}{\sum_C b_{oC\cdot}[1 - w(s, C)]},$$

where $b_{oCD} = \sum_{j \in J_o} b_{jCD}$, and $b_{oC\cdot} = \sum_D b_{oCD}$.

Following a similar development for the parameters λ_{os}, the equation for the adjustment of the estimates of such parameters turns out to be:

$$\lambda_{os} = \frac{\sum_C \sum_D b_{oCD} w(s, C)[1 - w(s, D)]}{\sum_C b_{oC\cdot} w(s, C)}.$$

Estimation of the careless error and lucky guess parameters

In each iteration of the EM algorithm, the function $U_3(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is maximized by setting to zero its first partial derivatives with respect to the parameters α_q and β_q. The first partial derivative of $U_3(\boldsymbol{\theta}, \boldsymbol{\theta}')$ with respect to the parameter α_q turns out to be:

$$\frac{\partial U_3(\boldsymbol{\theta}, \boldsymbol{\theta}')}{\partial \alpha_q} = \sum_{j=1}^{i} \sum_{C \subseteq S} \sum_{D \subseteq S} b_{jCD} \left[\frac{(1 - x_{1jq})v(q, C) + (1 - x_{2jq})v(q, D)}{\alpha_q} - \frac{x_{1jq}v(q, C) + x_{2jq}v(q, D)}{1 - \alpha_q} \right], \quad (2.9)$$

where $v(q, C) \in \{0, 1\}$ (resp. $v(q, D)$) is equal to 1 if problem q is solvable by state C (resp. D). By setting the right-hand term of (2.9) to zero, it follows that:

$$(1 - \alpha_q) \sum_j \sum_C \sum_D b_{jCD}[(1 - x_{1jq})v(q, C) + (1 - x_{2jq})v(q, D)] = \alpha_q \sum_j \sum_C \sum_D b_{jCD}[x_{1jq}v(q, C) + x_{2jq}v(q, D)]. \quad (2.10)$$

By solving (2.10) for α_q, one obtains the equation for the adjustment of the estimates of such parameters in each iteration of the EM algorithm:

$$\alpha_q = \frac{\sum_j \sum_C [b_{jC\cdot}(1 - x_{1jq}) + b_{j\cdot C}(1 - x_{2jq})]v(q, C)}{\sum_j \sum_C (b_{jC\cdot} + b_{j\cdot C})v(q, C)},$$

where $b_{jC\cdot} = \sum_D b_{jCD}$, and $b_{j\cdot C} \sum_D b_{jDC}$.

Following a similar development for the parameters β_q, the equation for the adjustment of the estimates of such parameters turns out to be:

$$\beta_q = \frac{\sum_j \sum_C (b_{jC\cdot}x_{1jq} + b_{j\cdot C}x_{2jq})[1 - v(q, C)]}{\sum_j \sum_C (b_{jC\cdot} + b_{j\cdot C})[1 - v(q, C)]}.$$

Computation of the initial probabilities, gain and loss probabilities for individual students

The model specifications presented so far concern the use of the GaLoM at the classroom level. However, once the model has been estimated and validated on a suitable sample of students, it can be used to obtain diagnostic information at the student level. The MATLAB code for computing initial probabilities, gain and loss probabilities for individual students is provided in Appendix A.2. Careless error and lucky guess probabilities for individual students are not given because they are more informative at the group level for detecting misspecifications of the model and noise of the data.

Individual student information can be obtained through a computation of the relevant conditional probabilities. To make things a little bit simpler, a single group of students is considered (e.g., all students receiving a specific learning object). An extension to more than one group is, however, straightforward.

The probability that $C \subseteq S$ is the initial competence state of a student j, given the student's initial response pattern $\mathbf{r}_{j1}$ and the model parameters $\boldsymbol{\theta} = (\alpha, \beta, \pi, \gamma, \lambda)$, is obtained by a straightforward application of Bayes' theorem:

$$P(\mathbf{C}_1 = C|\mathbf{R}_1 = \mathbf{r}_{j1}, \boldsymbol{\theta}) = \frac{P(\mathbf{R}_1 = \mathbf{r}_{j1}|\mathbf{C}_1 = C, \boldsymbol{\theta})P(\mathbf{C}_1 = C|\boldsymbol{\theta})}{\sum_{X \subseteq S} P(\mathbf{R}_1 = \mathbf{r}_{j1}|\mathbf{C}_1 = X, \boldsymbol{\theta})P(\mathbf{C}_1 = X|\boldsymbol{\theta})}, \quad (2.11)$$

where $P(\mathbf{C}_1 = C|\boldsymbol{\theta})$ is regarded as the probability that the initial knowledge state of a randomly sampled student is C.

Then, the probability that student j possesses skill $s \in S$ at the time of the pretest is simply the probability of being, at that time, in any of the states containing s:

$$\hat{\pi}_{js} = P(s \in \mathbf{C}_1|\mathbf{R}_1 = \mathbf{r}_{j1}, \boldsymbol{\theta}) = \sum_{C \in \mathcal{C}_s} P(\mathbf{C}_1 = C|\mathbf{R}_1 = \mathbf{r}_{j1}, \boldsymbol{\theta}), \quad (2.12)$$

where $\mathcal{C}_s = \{C \subseteq S : s \in C\}$ is the collection of all states containing s. By inserting (2.11) into (2.12) one obtains:

$$\hat{\pi}_{js} = \frac{\sum_{C \in \mathcal{C}_s} P(\mathbf{R}_1 = \mathbf{r}_{j1}|\mathbf{C}_1 = C, \boldsymbol{\theta})P(\mathbf{C}_1 = C|\boldsymbol{\theta})}{\sum_{X \subseteq S} P(\mathbf{R}_1 = \mathbf{r}_{j1}|\mathbf{C}_1 = X, \boldsymbol{\theta})P(\mathbf{C}_1 = X|\boldsymbol{\theta})}.$$

The gain and loss probabilities for each single student and each skill are now considered. The joint probability that the initial and final states of student j are, respectively, $\mathbf{C}_1 = C$ and $\mathbf{C}_2 = D (C, D \subseteq S)$ can be derived from the basic equation of the GaLoM (Equation 2.1). It suffices to recall that, for the specific assumptions of the model, the joint probability $P(\mathbf{R}_1, \mathbf{R}_2, \mathbf{C}_1, \mathbf{C}_2|\boldsymbol{\theta})$ can be factored as:

$$P(\mathbf{R}_1, \mathbf{R}_2, \mathbf{C}_1, \mathbf{C}_2|\boldsymbol{\theta}) = P(\mathbf{R}_1|\mathbf{C}_1, \boldsymbol{\theta})P(\mathbf{R}_2|\mathbf{C}_2, \boldsymbol{\theta})P(\mathbf{C}_2|\mathbf{C}_1, \boldsymbol{\theta})P(\mathbf{C}_1|\boldsymbol{\theta}).$$

Applying Bayes' theorem one obtains:

$$P(C, D|\mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta}) = \frac{P(\mathbf{r}_{j1}|C, \boldsymbol{\theta})P(\mathbf{r}_{j2}|D, \boldsymbol{\theta})P(D|C, \boldsymbol{\theta})P(C|\boldsymbol{\theta})}{\sum_{X \subseteq S} \sum_{Y \subseteq S} P(\mathbf{r}_{j1}|X, \boldsymbol{\theta})P(\mathbf{r}_{j2}|Y, \boldsymbol{\theta})P(Y|X, \boldsymbol{\theta})P(X|\boldsymbol{\theta})}.$$

This equation is at the basis of the computation of individual gain and loss probabilities. The gain probability concerns the acquisition of a new skill $s \in S$. It is defined as the conditional probability that s is in the final state $\mathbf{C}_2$, given that it was not in the initial state $\mathbf{C}_1$ of the student. Indicating with $\bar{\mathcal{C}}_s = 2^S \setminus \mathcal{C}_s$ the collection of all states not containing s,

$$P(s \notin \mathbf{C}_1, s \in \mathbf{C}_2|\mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta}) = \sum_{C \in \bar{\mathcal{C}}_s} \sum_{D \in \mathcal{C}_s} P(C, D|\mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta}),$$

from which one obtains the gain probability for student j and skill s:

$$\begin{aligned} \hat{\gamma}_{js} &= P(s \in \mathbf{C}_2|s \notin \mathbf{C}_1, \mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta}) \\ &= \frac{P(s \notin \mathbf{C}_1, s \in \mathbf{C}_2|\mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta})}{1 - P(s \in \mathbf{C}_1|\mathbf{r}_{j1}, \boldsymbol{\theta})} \\ &= \frac{1}{1 - \hat{\pi}_{js}} \sum_{C \in \bar{\mathcal{C}}_s} \sum_{D \in \mathcal{C}_s} P(C, D|\mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta}). \end{aligned}$$

A similar development leads to the loss probability:

$$\begin{aligned} \hat{\lambda}_{js} &= P(s \notin \mathbf{C}_2|s \in \mathbf{C}_1, \mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta}) \\ &= \frac{P(s \in \mathbf{C}_1, s \notin \mathbf{C}_2|\mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta})}{P(s \in \mathbf{C}_1|\mathbf{r}_{j1}, \boldsymbol{\theta})} \\ &= \frac{1}{\hat{\pi}_{js}} \sum_{C \in \mathcal{C}_s} \sum_{D \in \bar{\mathcal{C}}_s} P(C, D|\mathbf{r}_{j1}, \mathbf{r}_{j2}, \boldsymbol{\theta}). \end{aligned}$$

2.2 A simulation study

The simulation study tests the GaLoM with respect to model identifiability and goodness-of-recovery. In particular, considering different levels of information and noise in the data, and effects of the learning object on the skills, it investigates under which conditions the model parameters are uniquely determined and well-recovered.

2.2.1 Simulation of data sets and estimation of the models

Eight thousand random data sets were generated according to the GaLoM and 16 conditions. These conditions were produced by considering two ratios between the number of problems and underlying skills, four combinations of learning object effects in gaining and losing the skills, and two levels of noise in the data. With respect to the first, two collections with 10 and 20 problems were generated, and five skills were set to underlie both. Each problem of the two collections has been associated with its competencies according to the skill multimaps represented in Table 2.1. The two resulting performance structures contain 32 states, and were used to generate the data for the conditions with 10 and 20 problems. The learning object was set to highly affect both gain and loss of the skills (γ_{true} and $\lambda_{true} \geq .66$), to highly affect gain and poorly affect loss ($\gamma_{true} \geq .66; \lambda_{true} \leq .33$), to poorly affect gain and highly affect loss ($\gamma_{true} \leq .33; \lambda_{true} \geq .66$), to poorly affect both gain and loss (γ_{true} and $\lambda_{true} \leq .33$)[3]. The noise in the data was set to be low (α_{true} and $\beta_{true} \leq .1$) in one case, and medium (α_{true} and $\beta_{true} \leq .3$) in the other case.

The true initial probabilities of the skills π_{true} were set to be in the interval

[3]High loss parameters such as those considered in the study would be improbable in real educational settings. In fact, they describe a situation which would only occur when learning objects teach false knowledge. Regardless of their plausibility, it was decided to include these parameters in order to test the functioning of the model under extreme conditions.

Table 2.1: Competency Assignment in Collections with 10 and 20 Problems

10 Problems collection		20 Problems collection			
Problem	Competencies	Problem	Competencies	Problem	Competencies
1	$\{a\}$	1	$\{a\}$	11	$\{b,c,e\},\{a,b,c\}$
2	$\{b\}$	2	$\{b\}$	12	$\{a,b,c,d\},\{c,d,e\}$
3	$\{c\}$	3	$\{c\}$	13	$\{d,e\},\{a,b,d\}$
4	$\{d\}$	4	$\{d\}$	14	$\{b,c\}$
5	$\{a,b\},\{e\}$	5	$\{e\}$	15	$\{b,d\}$
6	$\{b,e\}$	6	$\{a,b\},\{e\}$	16	$\{b,c,d,e\}$
7	$\{c,e\}$	7	$\{b,e\}$	17	$\{a,c,d,e\}$
8	$\{c,d\}$	8	$\{c,e\},\{a,b,c\}$	18	$\{a,c,e\}$
9	$\{a,b,c\},\{b,c,e\}$	9	$\{c,d\}$	19	$\{a,c\}$
10	$\{a,b,c,d\},\{c,d,e\}$	10	$\{a,e\}$	20	$\{b,c,d\}$

Note. Letters from a to e refer to the five skills.

$[.1, .9]$. This choice was informed by a preliminary set of simulations where it was observed that, with very high or very low initial probabilities, it becomes difficult to estimate gain and loss probabilities. This occurs because, when the initial probability of a skill is very close to, for example, 1, the variance of such a skill in the sample is extremely small. Therefore, it is not easy to estimate the probability of gaining the skill in the few students who do not possess it in the pretest. However, it should be noted that, when the initial probability of a skill is very high, estimating gain probability of such a skill is negligible. The same thing stands for estimating loss probability of a skill when the initial probability is very low. The true model parameters were randomly generated according to the aforementioned constraints. Initial probabilities of the skills, and gain and loss parameters, were fixed across the two noise conditions, whereas they varied in the conditions concerning the number of problems and the learning object effect.

For each of the 16 conditions, 500 random data sets were simulated by using a parametric bootstrap (see, e.g., Langeheine, Pannekoek, & van de Pol, 1996; von Davier, 1997). The number of response patterns was set to $1,000$ for each data set.

For each of the 500×16 random data sets, the models were estimated which incorporated the performance structures used to generate the data.

2.2.2 Testing model identifiability and goodness-of-recovery

Model identifiability was tested in each condition. A rigorous test for identifiability can be accomplished by checking if the Hessian matrix is positive definite and the Jacobian matrix is non-singular (see, e.g., Arminger, Clogg, & Sobel, 1995). Such a strict analysis goes beyond the purposes of exploring the functioning of the GaLoM in different conditions. Model identifiability was tested in the following way. One of the 500 simulated data sets was randomly selected from each condition, and the model parameters were estimated 100 times, by randomly varying their initial values between 0 and 1. The α and β parameters were randomly generated between 0 and .5 because, when they are higher than .5, their interpretation as lucky guess and careless error is meaningless. A model was taken to be identifiable when the standard deviations were less than .01 for all the parameters.

The goodness-of-recovery was tested by considering the bias between the true parameters and the mean of the parameters estimates reproduced on the 500 simulated data sets.

2.2.3 Results

Concerning the problem of identifiability, the models are identifiable (i.e., $SD <$ .01 for all parameters) in all the conditions with 20 problems, and in the condition γ_{true} and $\lambda_{true} \leq .33$ with 10 problems. In the aforementioned conditions, the models are identifiable with both levels of noise in the data. On the contrary, at least one parameter has multiple solutions in the other conditions with 10 problems. This result suggests that model identifiability depends on the ratio between the number of problems and that of underlying skills, and on the learning object effect.

The following part concerns the goodness-of-recovery. Only the conditions in which the model is identifiable are considered. For the condition γ_{true} and

$\lambda_{true} \leq .33$ with 10 problems and noise $\leq .1$, Figure 2.1 depicts the true parameters (x axis) plotted versus the mean of the parameter estimates (and the related standard errors) reproduced on the 500 simulated data sets (y axis). The empirical biases of the estimates are negligible for most of the parameters. Standard errors of the estimates are also quite small. Figure 2.2 depicts the recovery of the model parameters for the same condition with noise $\leq .3$. It can be seen that empirical biases and standard errors of the estimates are larger than in the condition with noise $\leq .1$.

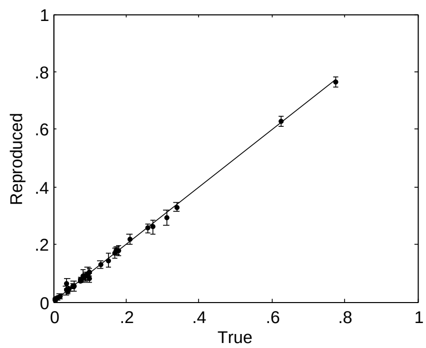

Figure 2.1: True parameters (x axis) versus mean of the parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis). Condition γ and $\lambda \leq .33$ with 10 problems and noise $\leq .1$. The straight line $x = y$ is added for reference.

Figure 2.3 depicts the four conditions with 20 problems and noise $\leq .1$. The empirical bias of the estimates is negligible for all the parameters when γ_{true} and $\lambda_{true} \geq .66$, $\gamma_{true} \geq .66$ and $\lambda_{true} \leq .33$, and γ_{true} and $\lambda_{true} \leq .33$, and it is rather high for a few parameters when $\gamma_{true} \leq .33$ and $\lambda_{true} \geq .66$.

A closer look at the latter condition highlights that there is a relationship among the parameters with the highest biases. These parameters are the γ and λ of skill a, and the α of problem 1, which is associated with skill a (see lower left diagram in Figure 2.3). Considering that skill a has the lowest initial probability ($\pi_{true} = .26$) and an extremely high loss probability ($\lambda_{true} = .97$), the cause of

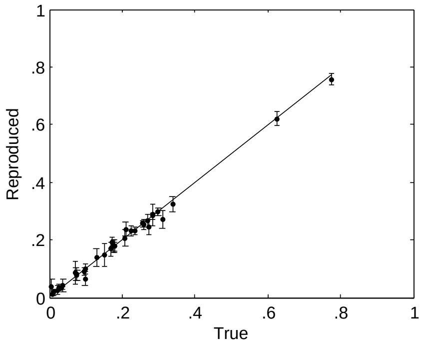

Figure 2.2: True parameters (x axis) versus mean of the parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis). Condition γ and $\lambda \leq .33$ with 10 problems and noise $\leq .3$. The straight line $x = y$ is added for reference.

the biases is understood. It is difficult for the GaLoM to reproduce the high loss of a skill when the initial probability of that skill is low. The underestimation of the loss probability of skill a ($\lambda_{rep} = .77$) is compensated by the overestimation of its gain probability ($\gamma_{true} = .03; \gamma_{rep} = .23$) and that of the careless error of problem 1 ($\alpha_{true} = .02; \alpha_{rep} = .34$). It is reasonable to hypothesize that the effect on problem 1 is stronger than that on the other problems that are associated with skill a because problem 1 is only associated with skill a. The estimates concerning the initial probability of skill a and the lucky guess of problem 1 are unbiased. Perhaps the biases in the present condition are not due to the learning object effect that is considered but to the specific values of the true parameters that were used for generating the data. In the other three conditions there are skills neither with small initial probabilities and very high loss probabilities nor with high initial probabilities and very high gain probabilities, and all the parameters are well-recovered.

Figure 2.4 depicts the four conditions with 20 problems and noise $\leq .3$. The empirical biases and standard errors of the estimates become larger than those of the conditions with noise $\leq .1$. This result suggests that the model is capable

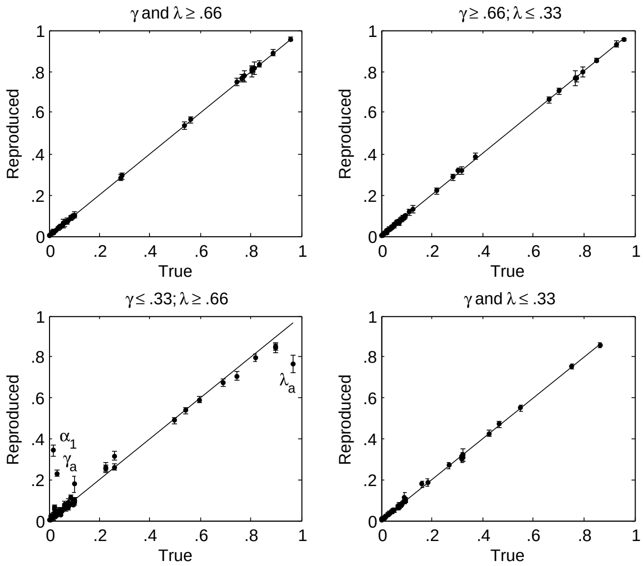

Figure 2.3: True parameters (x axis) versus mean of the parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis). Conditions with 20 problems and noise $\leq .1$. The straight line $x = y$ is added for reference.

of recovering the true parameters. Goodness-of-recovery improves when the gain and loss probabilities of the skills are not too high with respect to their initial probabilities, as well as when the noise in the data is low.

2.2.4 Discussion

The GaLoM was tested in a simulation study with respect to model identifiability and goodness-of-recovery. Different levels of information and noise in the data, and effects of the learning object on the skills have been considered.

Model identifiability seems to depend on the ratio between the number of problems and underlying skills, and on the learning object effect. It is reasonable to expect that data provide enough information about the skills so that stable and reliable estimates can be obtained. The simulations show that, when the skills are assessed by means of a small number of problems, compensations among

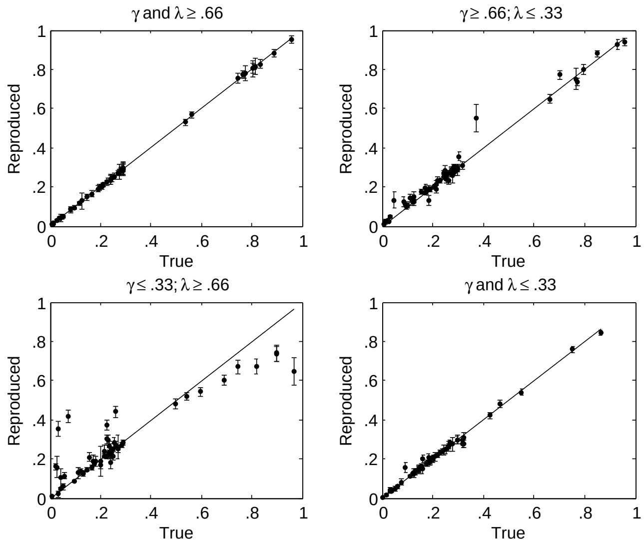

Figure 2.4: True parameters (x axis) versus mean of the parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis). Conditions with 20 problems and noise $\leq .3$. The straight line $x = y$ is added for reference.

model parameters are not excluded. The analysis of the parameters with multiple solutions highlights that there is a relation between the α and β probabilities of a problem and the π,γ, and λ probabilities of the skills that are associated with that problem. This is the case, for example, for condition γ_{true} and $\lambda_{true} \geq .66$ with 10 problems and noise $\leq .1$, in which the parameters that compensated each other were the π, γ and λ of skill a, and the α and β of problem 1, that was associated with that skill.

Among the conditions with 10 problems, γ_{true} and $\lambda_{true} \leq .33$ was the only one in which the model was identifiable. This condition describes a situation in which there is no change between the two assessment steps. Therefore, even if it does not provide much information, this information is consistent in the two assessment steps. This could be the reason for which it has been possible to obtain unique estimates of the parameters.

The level of information in the data is important for a model to be identifiable in the different conditions of learning object effect. Beyond the simple ratio between the number of problems and underlying skills, it is reasonable to expect that an important role in model identifiability is played by the way the problems and the skills are related to each other, that is, by the specification of the skill multimap. Moreover, it is reasonable to expect that the level of noise in the data may affect model identifiability, even if this has not been the case in the preset simulations.

In all cases with no identification problems, the model is able to recover the true parameters. Not surprisingly, recovery of parameters improves when the noise in the data is low. Goodness-of-recovery depends on the values of initial probabilities, and gain and loss probabilities. In fact, it is difficult for the GaLoM to reproduce a very high loss of a skill when the initial probability of that skill is low. It is reasonable to expect that the same thing holds for reproducing a very high gain when the initial probability is high. This result fits with the observation that it is difficult to estimate gain and loss probabilities when the initial probabilities are very high or very low.

2.3 An empirical application

The empirical application tests the capability of the GaLoM to assess the effect of learning objects on the attainment of specific skills. In particular, whether or not the model correctly distinguishes between an effective and an ineffective learning object is investigated. The analysis is performed with respect to the classroom and to individual students.

2.3.1 Method

Sixty-seven psychology students at the University of Padua participated in the study with no financial reward. The students were attending the course of Psychometrics in the academic year 2007-2008. Their mean age was 20.57 ($SD = 1.05$; range from 19 to 25), and 54 were female. A collection of 13 open response problems in elementary probability theory (see Appendix B.1) was presented through a computer-based testing procedure. The data were collected in a computer room that contained 40 workstations. Four skills (stochastic independence, law of total probability, conditional probability, probability of the complement of an event), and their combinations, were assumed to be required for solving the problems.

With the aim of testing the model, a 2×2 experimental design with two learning objects (effective vs. ineffective) and two assessment steps (pretest and posttest) was planned. The effective learning object was assumed to be useful for learning the skills required to solve the problems, whereas the ineffective learning object was assumed not to be useful. To give an example, consider the following problem (problem 5) taken from the collection:

> "Given two events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .86; P(A \cap \overline{B}) = .02$. Find $P(\overline{A})$".

It is assumed that the problem requires the skills concerning the probability of the complement of an event and the total probability to be solved. The effective learning object was the following:

> "If S is a sample space, then:
>
> $P(\overline{X}) = 1 - P(X)$ for any $X \subseteq S$ (for $X \subseteq S, \overline{X}$ is the complement of X in S).
>
> $P(X) = P(X \cap Y) + P(X \cap \overline{Y})$ for any $X, Y \subseteq S$.
>
> $P(X|Y) = P(X \cap Y)/P(Y)$ for any $X, Y \subseteq S$.
>
> $P(X \cap Y) = P(X)P(Y)$ if $X, Y \subseteq S$ are mutually independent",

and it was assumed to be useful for attaining the skills required to solve problem 5. The ineffective learning object was:

> "A finite stochastic process is a finite sequence of experiments in which each experiment has a finite number of outcomes with given probabilities.
>
> The probability of the empty set is 0.
>
> If S is a sample space, then: $P(X - Y) = P(X) - P(X \cap Y)$",

and it was assumed not to be useful for solving problem 5.

After responding to the problems the first time (pretest), students belonging to a first group (Group E, $N = 36$) were presented with the effective learning object, and those belonging to a second one (Group I, $N = 31$) with the ineffective learning object. Then, a posttest with the same problems took place. The responses to the problems were coded as correct (1) or incorrect (0).

2.3.2 Model estimation

In this application example, the assumption was made that the relation between problems and skills could have been described through the conjunctive model[4]. The conjunctive model is a special case of the competency model. In fact, it corresponds to a competency model in which each problem is associated only with one competency. Each problem from the collection was associated with the skills that were assumed to be necessary and sufficient for its mastery according to the skill map represented in Table 2.2. The performance structure delineated by the given skill map is a simple closure space containing 16 states. It is depicted in Figure 2.5.

[4]Via the conjunctive model, a skill map associates each problem with a subset of skills that are necessary and sufficient to solve it. A skill map is a triple (Q, S, σ), where Q is a non-empty set of problems, S is a non-empty set of skills, and σ is a mapping from Q to 2^S such that $\sigma(q) \neq \emptyset$ (i.e., each problem is associated with at least one skill). The performance state delineated by the subset of skills $X \subseteq S$ via the conjunctive model is specified by: $N(X) = \{q \in Q | \sigma(q) \subseteq X\}$, and the performance structure delineated by the skillmap (Q, S, σ) is: $\mathcal{K} = \{N(X) | X \subseteq S\}$ (Doignon, 1994; Doignon & Falmagne, 1999).

After estimating and validating the model at the classroom level, the probabilities that are relevant for the assessment of individual students were computed.

Table 2.2: Skill Assignment in the Conjunctive Model

Problem	Skills
1	{cp}
2	{tt}
3	{cd}
4	{id}
5	{cp, tt}
6	{cp, cd}
7	{cp, id}
8	{tt, cd}
9	{tt, id}
10	{cd, id}
11	{cp, tt, cd}
12	{cp, tt, id}
13	{cp, cd, id}

Note. cd = conditional probability; cp = complement of an event; id = stochastic independence; tt = total probability.

2.3.3 Testing model identifiability, goodness-of-fit and goodness-of-recovery

To test model identifiability, parameters were estimated 100 times, by randomly varying their initial values between 0 and 1.

The goodness-of-fit was evaluated using Pearson's Chi-square statistic. When the data sample is large enough, the statistic approximates the asymptotic Chi-square distribution well and such distribution can be used for statistic inference. In contrast, the approximation to the asymptotic Chi-square distribution of the statistic lacks accuracy for large and sparse data matrices. This is the case of the present study, because with 13 problems and two assessment steps the theoretical number of distinct binary response patterns is huge ($2^{13\times2}$), and the observed data

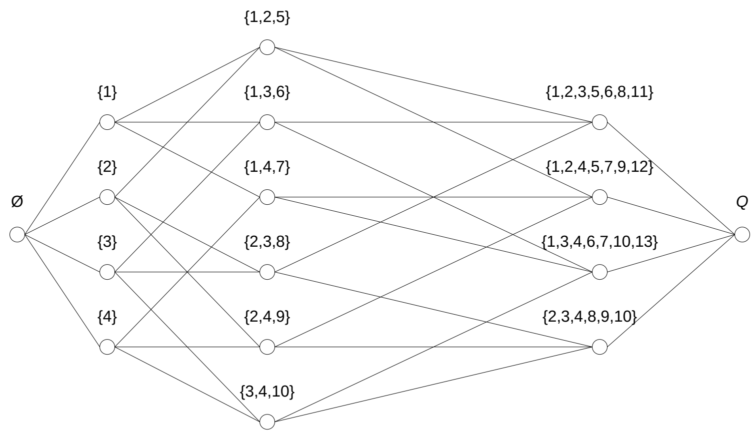

Figure 2.5: Hasse diagram of the performance structure delineated via the conjunctive model.

sample of 36 + 31 response patterns is definitely too small. A parametric bootstrap was therefore used for testing the model (see, e.g., Langeheine et al., 1996; von Davier, 1997). Using the parameters of the model estimated on the observed data, 1,000 random data samples of the same size of our sample (i.e., 36 + 31 response patterns) were simulated, and on these samples the model parameters were estimated. Then, the proportion of random data samples whose Chi-square was less than the Chi-square of the observed data sample was computed.

The goodness-of-recovery was tested by computing the bias between the parameter estimates obtained on the observed data sample and the mean of the parameter estimates reproduced on the 1,000 simulated data samples.

2.3.4 Results

Given that the proportion of random data samples whose Chi-square was less than the Chi-square of the observed data sample was .12, the goodness-of-fit of the model is good. Moreover, parameter estimates did not change in any of the 100 replications from different initial values, an indication that the model is identifiable.

With respect to the goodness-of-recovery, Figure 2.6 represents the parameter estimates obtained on the observed data sample (x axis) plotted versus the mean of the parameter estimates (and the related standard errors) reproduced on the 1,000 simulated data samples (y axis). Plots concerning loss and gain probabilities of Group I are not depicted because these probabilities are less than .1 for all the skills, both of them were estimated on the observed data sample and reproduced on the simulated ones. The empirical bias of the estimates is negligible for most of the parameters. Bootstrapped standard errors of the estimates are also quite small (from .01 to .19, see Tables 2.3 and 2.4).

Table 2.3 contains the estimates of the parameters α and β. The careless error probabilities are rather high for problems 9, 11, 12, and 13 ($\alpha_9 = .51; \alpha_{11} = .63; \alpha_{12} = .69; \alpha_{13} = .49$). Considering that the problems were open response, the lucky guess probabilities are quite high for problems 1, 3, and 13 ($\beta_1 = .31; \beta_3 = .35; \beta_{13} = .25$). This result suggests that the skill map has not been properly specified for these problems, even if the model fit is good.

Table 2.3: Maximum Likelihood Estimates of the Parameters α and β

	Careless error		Lucky guess			Careless error		Lucky guess	
Problem	α	SE	β	SE	Problem	α	SE	β	SE
1	.02	.02	.31	.10	8	.29	.10	.06	.03
2	.22	.06	.04	.04	9	.51	.09	.03	.02
3	.02	.03	.35	.06	10	.07	.07	.17	.04
4	$< .01$	.02	.20	.05	11	.63	.10	.02	.01
5	.09	.06	$< .01$	.01	12	.69	.09	$< .01$	.01
6	.04	.04	.13	.04	13	.49	.13	.25	.04
7	.25	.08	.02	.01					

Note. Standard errors (SE) of the estimates were obtained by parametric bootstrap.

Table 2.4 contains the maximum likelihood estimates of the parameters π, γ and λ. *Complement of an event* is the knowledge having the highest initial probability ($\pi_{cp} = .79$). *Total probability* ($\pi_{tt} = .49$), *conditional probability* ($\pi_{cd} = .36$), and *stochastic independence* ($\pi_{id} = .35$) follow.

The learning object presented to Group E has been effective in promoting the acquisition of knowledge. *Total probability* is the skill attained with the highest probability ($\gamma_{tt} = .80$), followed by *stochastic independence* ($\gamma_{id} = .48$) and *conditional probability* ($\gamma_{cd} = .23$). The skill concerning the *complement of an event* is not further acquired ($\gamma_{cp} < .01$). Given that it is already the skill with the highest initial probability, it is difficult to find the effectiveness of the learning object in the few students who do not yet possess it. Unlike the learning object presented to group E, the one presented to Group I has not been effective on the attainment of the skills ($\gamma < .01$ for all parameters).

Unexpectedly, in Group E the probability of losing three of the four skills is greater than in Group I, even if these probabilities are quite small (see Table 2.4). Perhaps, this result is due to a compensation effect between parameters γ and λ that is observed only in Group E because in this group learning occurs between the two assessment steps.

Table 2.4: Maximum Likelihood Estimates of the Parameters π, γ and λ

			Group E ($N = 36$)				Group I ($N = 31$)			
	Initial p.		Gain		Loss		Gain		Loss	
Skill	π	SE	γ	SE	λ	SE	γ	SE	λ	SE
Complement of an event	.79	.06	$< .01$	.05	$< .01$	.01	$< .01$	.03	$< .01$	.02
Total probability	.49	.08	.80	.19	.09	.08	$< .01$	.01	$< .01$	.02
Conditional probability	.36	.07	.23	.10	.02	.06	$< .01$	.02	$< .01$	.03
Stochastic independence	.35	.07	.48	.13	.10	.10	$< .01$	.01	$< .01$	.02

Note. Standard errors (SE) of the estimates were obtained by parametric bootstrap. Group E = effective learning object; Group I = ineffective learning object.

The probability of skill s at the end of the learning process as a result of learning object o that has been presented can be easily calculated as $\pi_s(1 - \lambda_{os}) + (1 - \pi_s)\gamma_{os}$. This probability will be called final probability. By way of example, the probability of the skill concerning the total probability increases from .49 to .85 in the group with the effective learning object, and remains .49 in the group with the ineffective learning object.

The probability of the competence states at the pretest can be computed according to Equation (2.2). Similarly, we can use the final probabilities of the skills for computing the probability of the competence states at the posttest. Figure 2.7 depicts the probability distributions of the competence states at the pretest (computed on the whole sample of students) and at the posttest (computed separately for Groups E and I). The competence state containing *complement of an event*, and that containing *complement of an event* and *total probability* are the most probable competence states at the pretest. Compared with the pretest, in Group E the probabilities of the competence states containing more skills increase at the posttest, whereas those of the competence states containing less skills decrease. In particular, the competence state containing *complement of an event*, *total probability* and *stochastic independence*, and that containing all the skills are the most probable competence states at the posttest. On the contrary, in Group I the probabilities of the competence states do not change between the pretest and the posttest. Concerning the acquisition of knowledge, this result suggests that the model correctly distinguishes between the effective learning object and the ineffective one.

Tables 2.5 and 2.6 contain initial and final probabilities of the skills, and gain and loss probabilities for each student of Group E and Group I, respectively. For most of the students of Group E the gain probabilities of the skills are greater than the loss probabilities and, as expected, the final probabilities of the skills are greater than the initial ones. In contrast, for students of Group I the initial and final probabilities of the skills do not differ from each other. The tables also contain the competence state which is modal for each student. For most students of Group E the competence state on the posttest contains more skills than that on the pretest. This does not happen for the students of Group I. As expected, the learning object presented to Group E has a positive effect on the learning of the four skills. This is not the case of the learning object presented to Group I.

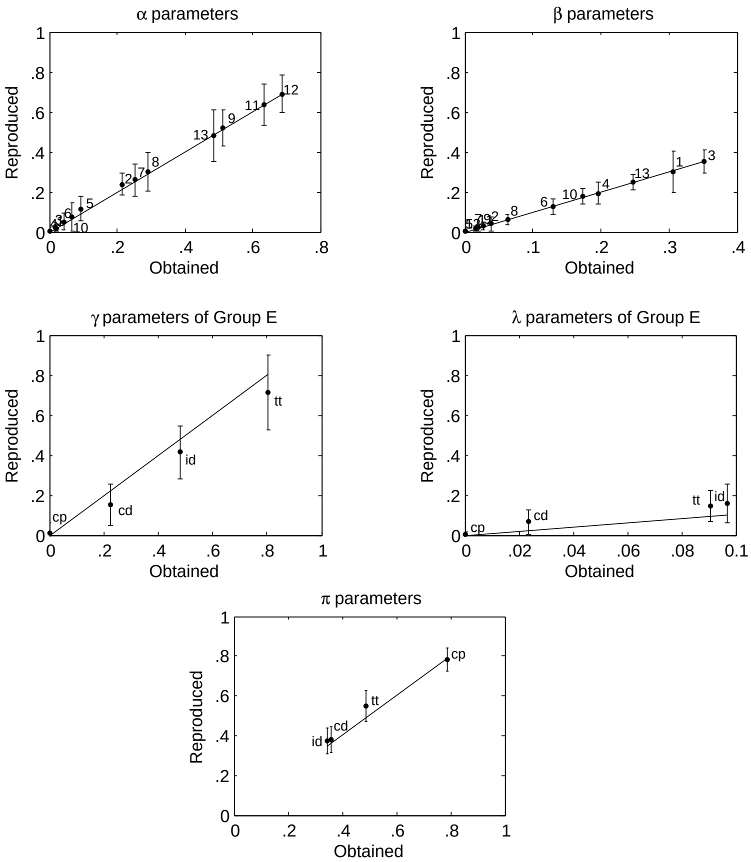

Figure 2.6: Parameter estimates obtained on the observed data (x axis) versus mean parameter estimates (and related standard errors) reproduced on the simulated data (y axis). Numbers in the plots refer to the problems (from 1 to 13). Group E = effective learning object; cd = conditional probability; cp = complement of an event; id = stochastic independence; tt = total probability. The straight line $x = y$ is added for reference.

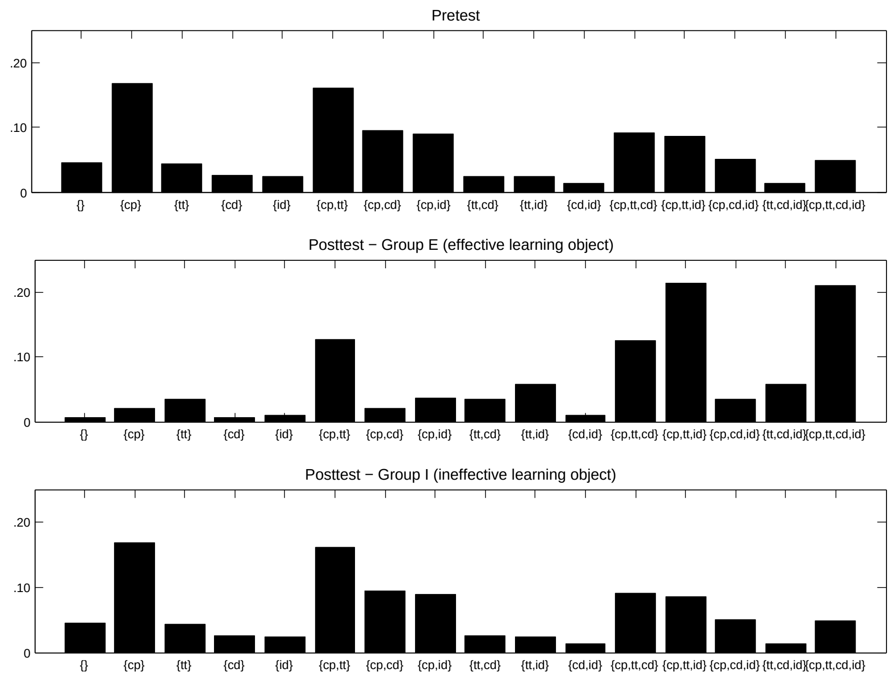

Figure 2.7: Probability distributions of competence states at the pretest (upper diagram) and at the posttest (central diagram for Group E, lower diagram for Group I). cd = conditional probability; cp = complement of an event; id = stochastic independence; tt = total probability.

Table 2.5: Initial and Final Probabilities of the Skills, Gain and Loss Probabilities, Initial and Final Competence States for Each Student of Group E (Effective Learning Object)

	Complement of an event				Total probability				Conditional probability				Stochastic independence				Competence state	
Student	Initial prob.	Final prob.	Gain	Loss	Initial prob.	Final prob.	Gain	Loss	Initial prob.	Final prob.	Gain	Loss	Initial prob.	Final prob.	Gain	Loss	Initial	Final
1	.99	1.00	1.00	< .01	.57	1.00	1.00	< .01	.99	1.00	1.00	< .01	.99	1.00	.99	< .01	{cp, tt, cd, id}	{cp, tt, cd, id}
2	1.00	1.00	a	< .01	.99	1.00	1.00	< .01	.99	1.00	.98	< .01	< .01	1.00	1.00	< .01	{cp, tt, cd}	{cp, tt, cd, id}
3	.70	.70	< .01	< .01	.13	.03	.02	.90	.07	.06	.01	.25	.67	.58	.20	.23	{cp, id}	{cp, id}
4	< .01	< .01	< .01	< .01	.50	.15	.09	.80	< .01	< .01	< .01	.91	< .01	< .01	< .01	1.00	{tt}	{}
5	1.00	1.00	a	< .01	.99	1.00	1.00	< .01	.99	1.00	.94	< .01	.99	1.00	1.00	< .01	{cp, tt, cd, id}	{cp, tt, cd, id}
6	1.00	1.00	a	< .01	.99	1.00	1.00	< .01	.99	.98	.31	.02	.99	1.00	.98	< .01	{cp, tt, cd, id}	{cp, tt. cd, id}
7	.99	1.00	1.00	< .01	.02	1.00	1.00	< .01	.96	1.00	1.00	< .01	.98	1.00	1.00	< .01	{cp, cd, id}	{cp, tt, cd, id}
8	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	< .01	.21	.20	.03	< .01	.99	.99	< .01	{cp}	{cp, tt, id}
9	.99	1.00	1.00	< .01	.01	1.00	1.00	< .01	.99	1.00	.94	< .01	.99	1.00	1.00	< .01	{cp, cd, id}	{cp, tt, cd, id}
10	.99	1.00	.99	< .01	.05	1.00	1.00	< .01	.94	.76	.03	.19	< .01	< .01	< .01	1.00	{cp, cd}	{cp, tt, cd}
11	.01	.01	< .01	< .01	.57	.92	.88	.05	< .01	.03	.03	.19	.63	< .01	< .01	1.00	{tt, id}	{tt}
12	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	.02	.98	.98	< .01	< .01	1.00	1.00	< .01	{cp}	{cp, tt, cd, id}
13	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	.02	.93	.93	< .01	< .01	.90	.90	.01	{cp}	{cp, tt, cd, id}
14	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	< .01	< .01	< .01	.78	< .01	.08	.08	.52	{cp}	{cp, tt}
15	.99	1.00	1.00	< .01	.09	1.00	1.00	< .01	< .01	.24	.24	.02	.82	.94	.78	.03	{cp, id}	{cp, tt, id}
16	.99	1.00	< .01	< .01	.02	.07	.07	.84	.99	.99	.39	.01	< .01	< .01	< .01	1.00	{cp, cd}	{cp, cd}
17	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	.02	.94	.94	< .01	< .01	.98	.98	< .01	{cp}	{cp, tt, cd, id}
18	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	.02	< .01	< .01	.97	< .01	.94	.94	.01	{cp}	{cp, tt, id}
19	.02	.02	< .01	< .01	.66	.91	.86	.06	< .01	< .01	< .01	.98	< .01	.34	.34	.16	{tt}	{tt}
20	.99	1.00	.88	< .01	.99	1.00	1.00	< .01	< .01	.05	.05	.13	< .01	.74	.74	.03	{cp, tt}	{cp, tt, id}
21	.99	1.00	1.00	< .01	.01	1.00	1.00	< .01	.99	.99	.58	.01	.99	1.00	.99	< .01	{cp, cd, id}	{cp, tt, cd, id}
22	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	< .01	.04	.04	.16	< .01	.09	.09	.51	{cp}	{cp, tt}
23	.99	1.00	1.00	< .01	.92	1.00	1.00	< .01	< .01	< .01	< .01	.87	.31	.20	.08	.53	{cp, tt}	{cp, tt}
24	.01	.01	< .01	< .01	.53	.89	.84	.07	< .01	.01	.01	.31	.87	.77	.34	.16	{tt, id}	{tt, id}
25	.99	1.00	< .01	< .01	.99	.99	< .01	< .01	.99	.99	.58	.01	.99	1.00	.97	< .01	{cp, tt, cd, id}	{cp, tt, cd, id}
26	.98	.98	< .01	< .01	.11	.27	.24	.51	< .01	.04	.04	.17	< .01	.95	.95	.01	{cp}	{cp, id}
27	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	.02	.81	.81	< .01	< .01	.02	.02	.83	{cp}	{cp, tt, cd}
28	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	< .01	.01	.01	.5	< .01	1.00	1.00	< .01	{cp}	{cp, tt, id}
29	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	.02	.98	.98	< .01	< .01	1.00	1.00	< .01	{cp}	{cp, tt, cd, id}
30	.99	1.00	1.00	< .01	.01	1.00	1.00	< .01	.99	1.00	.98	< .01	.99	1.00	1.00	< .01	{cp, cd, id}	{cp, tt, cd, id}
31	.16	.16	< .01	< .01	.31	.94	.93	.01	.76	.95	.79	< .01	.59	.81	.62	.06	{cd, id}	{tt, cd, id}
32	.99	1.00	1.00	< .01	.99	1.00	1.00	< .01	< .01	< .01	< .01	1.00	< .01	< .01	< .01	1.00	{cp, tt}	{cp, tt}
33	.99	1.00	1.00	< .01	.17	1.00	1.00	< .01	< .01	.04	.04	.16	< .01	.09	.09	.51	{cp}	{cp, tt}
34	.95	.95	< .01	< .01	.10	.02	.01	.93	< .01	< .01	< .01	.96	< .01	< .01	< .01	1.00	{cp}	{cp}
35	< .01	< .01	< .01	< .01	.50	.15	.09	.80	< .01	< .01	< .01	.91	< .01	< .01	< .01	1.00	{tt}	{}
36	.99	1.00	1.00	< .01	.16	1.00	1.00	< .01	.05	.05	.01	.34	< .01	.92	.92	.01	{cp}	{cp, tt, id}

Note. [a]When the initial probability of the skill is exactly 1, computing an individual gain probability is meaningless.
cd = conditional probability; cp = complement of an event; id = stochastic independence; tt = total probability.

Table 2.6: Initial and Final Probabilities of the Skills, Gain and Loss Probabilities, Initial and Final Competence States for Each Student of Group I (Ineffective Learning Object)

	Complement of an event				Total probability				Conditional probability				Stochastic independence				Competence state	
Student	Initial prob.	Final prob.	Gain	Loss	Initial prob.	Final prob.	Gain	Loss	Initial prob.	Final prob.	Gain	Loss	Initial prob.	Final prob.	Gain	Loss	Initial	Final
1	1.00	1.00	a	< .01	1.00	1.00	a	< .01	.91	.91	< .01	< .01	< .01	< .01	< .01	.92	{cp, tt, cd}	{cp, tt, cd}
2	.99	.99	< .01	< .01	.99	.99	< .01	< .01	.99	.99	.09	< .01	< .01	< .01	< .01	.06	{cp, tt, cd}	{cp, tt, cd}
3	.99	.99	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	.99	.99	< .01	< .01	{cp, id}	{cp, id}
4	1.00	1.00	a	< .01	1.00	1.00	a	< .01	.99	1.00	.52	< .01	1.00	1.00	a	< .01	{cp, tt, cd, id}	{cp, tt, cd, id}
5	< .01	< .01	< .01	< .01	.05	.05	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	.44	{}	{}
6	.01	.01	< .01	< .01	.01	.01	< .01	< .01	.67	.67	< .01	< .01	.88	.88	< .01	< .01	{cd, id}	{cd, id}
7	1.00	1.00	a	< .01	1.00	1.00	a	< .01	< .01	< .01	< .01	< .01	.99	.99	< .01	< .01	{cp, tt, id}	{cp, tt, id}
8	1.00	1.00	a	< .01	1.00	1.00	a	< .01	.99	.99	.04	< .01	1.00	1.00	a	< .01	{cp, tt, cd, id}	{cp, tt, cd, id}
9	.99	.99	< .01	< .01	< .01	< .01	< .01	< .01	.66	.66	< .01	< .01	< .01	< .01	< .01	.72	{cp, cd}	{cp, cd}
10	1.00	1.00	a	< .01	1.00	1.00	a	< .01	< .01	< .01	< .01	.19	< .01	< .01	< .01	.89	{cp. tt}	{cp, tt}
11	< .01	< .01	< .01	< .01	.05	.05	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	.44	{}	{}
12	.06	.06	< .01	< .01	.76	.76	< .01	< .01	.01	.01	< .01	< .01	< .01	< .01	< .01	.53	{tt}	{tt}
13	1.00	1.00	a	< .01	.99	.99	.04	< .01	.91	.91	< .01	< .01	< .01	< .01	< .01	.92	{cp, tt, cd}	{cp, tt, cd}
14	.99	.99	< .01	< .01	.99	.99	< .01	< .01	< .01	< .01	< .01	.19	< .01	< .01	< .01	.89	{cp, tt}	{cp, tt}
15	< .01	< .01	< .01	< .01	.99	.99	< .01	< .01	< .01	< .01	< .01	.01	< .01	< .01	< .01	.59	{tt}	{tt}
16	< .01	< .01	< .01	< .01	.99	.99	< .01	< .01	< .01	< .01	< .01	.01	< .01	< .01	< .01	.59	{tt}	{tt}
17	< .01	< .01	< .01	< .01	.05	.05	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	.44	{}	{}
18	1.00	1.00	a	< .01	1.00	1.00	a	< .01	< .01	< .01	< .01	.19	< .01	< .01	< .01	.89	{cp, tt}	{cp, tt}
19	.98	.98	< .01	< .01	< .01	< .01	< .01	< .01	.99	.99	< .01	< .01	.92	.92	< .01	< .01	{cp, cd, id}	{cp, cd, id}
20	.99	.99	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	.99	.99	< .01	< .01	{cp, id}	{cp, id}
21	.99	.99	< .01	< .01	< .01	< .01	< .01	< .01	.99	.99	.01	< .01	.99	.99	< .01	< .01	{cp, cd, id}	{cp, cd, id}
22	1.00	1.00	a	< .01	1.00	1.00	a	< .01	.99	.99	.09	< .01	< .01	< .01	< .01	.42	{cp, tt, cd}	{cp, tt, cd}
23	1.00	1.00	a	< .01	1.00	1.00	a	< .01	< .01	< .01	< .01	.04	.81	.81	< .01	< .01	{cp, tt, id}	{cp, tt, id}
24	.99	.99	< .01	< .01	< .01	< .01	< .01	< .01	.99	.99	< .01	< .01	< .01	< .01	< .01	.95	{cp, cd}	{cp, cd}
25	1.00	1.00	a	< .01	1.00	1.00	a	< .01	< .01	< .01	< .01	.19	.11	.11	< .01	< .01	{cp, tt}	{cp, tt}
26	.99	.99	< .01	< .01	.99	.99	< .01	< .01	< .01	< .01	< .01	< .01	.11	.11	< .01	< .01	{cp, tt}	{cp, tt}
27	1.00	1.00	a	< .01	1.00	1.00	a	< .01	.99	.99	.25	< .01	.99	1.00	1.00	< .01	{cp, tt, cd. id}	{cp, tt, cd, id}
28	.99	.99	< .01	< .01	.99	.99	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01	.89	{cp, tt}	{cp, tt}
29	1.00	1	a	< .01	1	1	a	< .01	.99	.99	.03	< .01	.99	.99	< .01	< .01	{cp, tt, cd, id}	{cp, tt, cd, id}
30	< .01	< .01	< .01	< .01	.81	.81	< .01	< .01	.01	.01	< .01	< .01	< .01	< .01	< .01	.53	{tt}	{tt}
31	.78	.78	< .01	< .01	< .01	< .01	< .01	< .01	.06	.06	< .01	< .01	.53	.53	< .01	< .01	{cp}	{cp}

Note. [a]When the initial probability of the skill is exactly 1, computing an individual gain probability is meaningless.
cd = conditional probability; cp = complement of an event; id = stochastic independence; tt = total probability.

2.3.5 Discussion

The empirical application tested the capability of the GaLoM to distinguish between an effective and an ineffective learning object. The model was applied to the responses provided by two groups of university students to a collection of problems in elementary probability theory. Students of a group were presented with an effective learning object, students of the other group with an ineffective learning object. Four skills were assumed to underlie the responses to the problems.

The results show that the model correctly distinguished between the two learning objects. On the whole, the gain probabilities and the final probabilities of the skills were greater in the group presented with the effective learning object than in the other group. Moreover, in the group presented with the effective learning object, the competence states that were more likely at the posttest contained more skills than those that were more likely at the pretest. Analogous results were observed on individual students.

An explicit assumption was made about what skills were required for solving the problems. This assumption has been confirmed by model fit. However, the careless error and lucky guess probabilities of some problems were rather high, and this might be a sign that the skill assignments concerning these problems were not entirely correct. A high careless error could suggest that a problem requires more skills than expected, or that it has been associated with superfluous solution strategies. In the present case, the problems with higher careless errors required the combination of two or three skills to be solved. A higher-order skill could be introduced in the model accounting for the capability of the students to properly combine the basic skills in order to solve the problems. A high lucky guess could suggest that a problem has been associated with more skills than needed, or that there is an alternative solution strategy that has not been mapped. Problem 13 is given as an example of the latter: "Given two independent events A and B

in a sample space S, the following probability is known: $P(A|B) = .02$. Find $P(\overline{A})$". The lucky guess parameter of the problem is quite high for an open response problem ($\beta_{13} = .25$). According to the specified skill map, the problem requires the knowledge of three of the four skills to be solved (see Table 2.2). Nevertheless, there might be students who do not possess the required skills, and solve the problem anyway by simply subtracting the only probability value from 1. This possibility could be taken into account by introducing an alternative solution strategy for problem 13. The aforementioned possibilities for improving the model have not been pursued because they entail the introduction of additional skills and thus of additional parameters, and the number of problems which are available would not be sufficient for obtaining stable and trustworthy estimates.

2.4 Final remarks

A formal model for assessing knowledge and learning processes has been presented. Practical benefits for teaching and learning that derive from using the model are now explored and discussed.

The model focuses on the specific skills that a student must possess in order to solve the problems. This makes the model particularly suitable for didactic practice, because it enables the teacher to theoretically explain the observed responses, and to predict responses on another collection of problems. Moreover, the teacher is helped in identifying which skills should be taught so that a previously unsolvable problem becomes solvable.

The model provides diagnostic information that helps the teacher in planning the didactic interventions and in evaluating their effectiveness. This information can be obtained at both classroom and student levels.

At the classroom level, initial probabilities of the skills enable the teacher to identify what the classroom already knows and what it is ready to learn. Gain and loss parameters provide program evaluative information. The teacher is informed about the effectiveness of the didactic intervention that has been carried out, and is enabled to compare different didactic interventions by measuring their effect on the acquisition of specific skills.

Considering that gain and loss parameters are indexed by learning object and skill, they enable the teacher to select the best learning objects for the specific weaknesses of the classroom. In general, these learning objects should have both high gain probabilities and small loss probabilities of the skills. However, the relative importance of gain and loss parameters depends on the specific context the model is applied to. For example, loss parameters are useful in educational settings in which the teacher might be interested in assessing the loss of the skills a certain period after the didactic intervention was carried out. On the contrary, in applications such as the one presented, in which the posttest occurs immediately after the presentation of the learning object, loss estimates are not very important. As a consequence, they could be constrained to be equal to zero, reducing the number of parameters that have to be estimated. If the goodness-of-fit of the constrained model is not significantly worse than that of the unconstrained model, the first one will be preferred because it is more parsimonious. In another context, such as the treatment of senile dementia, loss parameters may be more important than gain parameters because the specialist would be interested mainly in selecting the interventions that are associated with the smallest loss of skills of the patients over time. In this case, the gain estimates could be constrained to be equal to zero.

Careless error and lucky guess parameters provide information concerning the goodness of the relationship between latent competence states and observable responses. When they are small, the teacher is informed that the skill multimap has been specified appropriately and that the collection of problems has been effective in assessing the knowledge of the students. This information enables the teacher to validate his hypothesis on the skills needed for solving the problems and his assessment instruments. High careless error and lucky guess parameters could be a sign of failings in the specification of the skill multimap or of noise in the data. With respect to the specification of the skill multimap, a high careless error could suggest that a problem requires more skills than expected, or that it has been associated with a superfluous solution strategy. Conversely, a high lucky guess could suggest that a problem has been associated with more skills than needed, or that there is an alternative solution strategy that has not been mapped. With respect to noise in the data, a high careless error could suggest fatigue, motivational decrease, or that the wording of the problem confuses the students leading them to fail it. Conversely, a high lucky guess rate could indicate that the wording of the problem suggests the correct response to the students. High careless error and lucky guess parameters are difficult to interpret in practical applications because both the skill multimap and the noise underlying the data are unknown. In the following two chapters, an extension of the GaLoM is described that can be useful for identifying the best skill multimap among a number of alternatives.

Once the model has been estimated and validated on a suitable sample of students, it can be used to obtain diagnostic information at the student level. The initial probabilities of the skills, and the gain and loss probabilities can be obtained for each student. Single student proficiencies can be evaluated through a detailed skill profile. Such a profile can be obtained both before and after a specific didactic intervention takes place. The latent set of skills possessed by a student can be recovered by taking as the competence state of the student the one that has the highest posterior probability, given the observed response pattern and the parameters of the model. Moreover, since the GaLoM is capable of comparing the effects of different didactic interventions on the acquisition of specific skills, it enables the teacher to select the best educational intervention for the specific weaknesses of each student.

The teacher is thus supported along two parallel pathways. Along the first pathway the outcome is a detailed skill profile of a single student, which is by itself a valuable tool for assessing what the student already knows and what that student is ready to learn next. The second pathway provides the teacher with substantive information on the effectiveness of the didactic tools to be used. As far as a single didactic tool is concerned, this information is twofold, as it attempts to answer the following two questions: (a) "Is this tool appropriate (has a low loss probability) for improving or consolidating a skill that this student already possesses?", and (b) "Is this tool appropriate (has a high gain probability) for facilitating the learning of a new specific skill?" By having available a whole collection of didactic tools in the form of exercises and instructions, these two questions can be restated as: (a) "Which tool is best (has the lowest loss probability) for consolidating a skill?", and (b) "Which tool is best (has the highest gain probability) for learning a new skill?" The teacher would thus be provided with an objective criterion for choosing didactic tools throughout the teaching of the course, and this choice could even be personalized according to the specific students' needs and skill profiles.

It should be noted that, once the model has been estimated and validated on a suitable sample of students, it can be used for individual assessment of new students without having to reestimate the model parameters.

The GaLoM is now compared with some formal models for the assessment of knowledge that are present in literature. The BLIM (Falmagne & Doignon, 1988a) is considered first, which is the basic model in knowledge space theory and has been the foundation of several approaches. This model has been applied in a number of contexts (see, e.g., Falmagne et al., 1990; Stefanutti, 2006; Taagepera et al., 1997; see also Doignon & Falmagne, 1999). Unlike the BLIM, which focuses on the solution behaviour at the level of problems, the GaLoM focuses on the discrete skills that underlie the responses of the problems. Moreover, the BLIM only assesses the knowledge of students, whereas the GaLoM also provides information about the effect of an educational intervention on promoting learning.

The Deterministic Inputs, Noisy AND gate (DINA) model (Haberman, 1979;

Junker & Sijtsma, 2001; Macready & Dayton, 1977) is a very common approach within the cognitive diagnosis framework. Applications have been described in different contexts (see, e.g., de la Torre, 2008; de la Torre & Douglas, 2004). Similarly to the GaLoM, it assumes the existence of discrete skills that underlie the responses to the problems. However, whereas the DINA is a conjunctive model, the GaLoM is a competency model based on a skill multimap. An extension of the DINA model has been recently proposed, which allows the consideration of multiple strategies of solving the problems (de la Torre & Douglas, 2008). Unlike the DINA, which only allows the definition of student skill profiles, the GaLoM allows the assessment of the effect of an educational intervention on the attainment of specific skills.

The Stochastic Learning Path (SLP) model (Falmagne, 1993; see also Falmagne, 1989, 1996; Falmagne & Lakshminarayan, 1994) is an approach proposed within the knowledge space theory framework which describes the progress of students learning in a particular field. Applications of the model can be found in Falmagne et al. (1990), and Lakshminarayan and Gilson (1998). As for the BLIM, from which it derives, the SLP is developed at the level of the problems without assuming underlying skills. The SLP assesses learning by analyzing the progress of students in a collection of learning paths (i.e., chains of performance states beginning with the empty state $\emptyset$ and finishing with the full state Q). In contrast, in the GaLoM the particular learning paths followed by the students in between the two assessment steps are not taken into account. Moreover, in the GaLoM there is the possibility of modelling a certain kind of forgetting through the loss parameters, whereas the SLP is only suitable for modelling monotonic learning processes.

A model has been recently proposed (Heller, Levene, Keenoy, Albert, & Hockemeyer, 2007) that describes the learning of students while they navigate through a learning environment. Similarly to the SLP, this model describes how students move along learning paths, but it also takes into account the skills and competencies underlying the problems. In this model the performance structure is derived from a prerequisite map associating each problem with a collection of subsets of learning objects that provide the content sufficient for solving it. In contrast, in the GaLoM the performance structure is derived from a skill multimap. Unlike the GaLoM, in order to be applicable this model requires the assumption that there is no loss of the skills.

Chapter 3

Parameter identifiability and recovery of the correct skill assignment in the Constrained Gain-Loss Model

The GaLoM is a formal model for assessing learning processes. It provides detailed information about the knowledge of the students and the effectiveness of the didactic interventions in promoting specific learning. This information helps the teacher in planning teaching and learning.

Two elements are needed to reach a trustworthy assessment of students and didactic interventions. On one hand, data are required that provide enough information about the skills and that are not too noisy. When this is not the case, the results of the assessment might be inaccurate. Moreover, compensations between the α and β probabilities of some problems and the π,γ, and λ probabilities of the skills which are associated with those problems might result in multiple solutions for their estimates.

On the other hand, information is required about which skills are measured by the assessment instrument and how they are related to the problems. In the GaLoM this information, which contributes to the construct validity of the assessment instrument (Messick, 1989, 1995), is translated into a skill multimap associating each problem with a collection of subsets of skills which are necessary and sufficient to solve it. The skill multimap is the core element that determines the quality of the assessment. If it is correctly specified, one would expect the error probabilities to be small for all problems. If they were high for some problems, then this would be a sign of misspecification of the skill multimap with respect to these problems. In particular, a high careless error could suggest that a problem requires more skills than expected, or that it has been associated with a superfluous solution strategy. On the contrary, a high lucky guess could suggest that the problem has been associated with more skills than needed, or that there is an alternative solution strategy that has not been mapped.

A systematic investigation of the effects of misspecifications of the skills assigned to a problem has been provided by Rupp and Templin (2008) for the DINA model (de la Torre & Douglas, 2004; Macready & Dayton, 1977). As already mentioned, the DINA is a conjunctive model for the assessment of knowledge. The authors found that, when a relevant skill is deleted from a problem, the α parameter is overestimated whereas the β parameter estimate remains accurate. On the contrary, when a superfluous skill is added to a problem, the β parameter is overestimated whereas the α parameter estimate remains accurate. In an extension of the DINA model that allows for multiple solution strategies, de la Torre and Douglas (2008) found that the omission of a relevant solution strategy from a problem causes an overestimation of its lucky guess probability. On the contrary, the inclusion of an irrelevant solution strategy to a problem causes an overestimation of its careless error probability. The misspecifications have predominantly local effects in the sense that they mostly affect the α and β estimates of the problems whose skill assignment is not correct. Moreover, they can affect the assessment of knowledge. In particular, Rupp and Templin (2008) pointed out that the complete deletion of certain skill combinations highly compromises the assessment of the students with such combinations in their skill profile.

Both the identifiability of the parameters and the recognizability of the skill multimap underlying the data are addressed in this chapter. Given that there is a relation between the α and β probabilities of the problems and the π, γ, and λ probabilities of the skills which are associated with them, one could expect that constraining the estimates of the α and β parameters only allows the reduction of the variability of the whole set of parameters in non-identifiable models. This procedure could also help to distinguish the skill multimap that is correctly specified from one that is not. Since misspecification of the skill multimap causes an overestimation of the α and β parameters, constraining their estimates should constrain the model incorporating an incorrect skill multimap more than the model incorporating the correct skill multimap. As a consequence, the fit of the former is expected to get worse to a great extent than that of the latter.

The log-barrier method is used to constrain the estimates of the α and β parameters to be less than or equal to an upper bound. The constrained version of the GaLoM is called Constrained Gain-Loss Model (CoGaLoM, Anselmi, Stefanutti, & Robusto, Submitted).

The chapter is organized as follows. In the next paragraph, the log-barrier method and the way it has been used for constraining the α and β parameters of the GaLoM are described. Then, a simulation study that tests the CoGaLoM with respect to the identifiability of the parameters and the recognizability of the skills assignment underlying the data is presented. The chapter concludes with some remarks concerning the proposed approach and the comparison of models.

3.1 The Constrained Gain-Loss Model

The log-barrier method (see, e.g., Fiacco & McCormick, 1990; Wright, 1997) is used for constraining the estimates of the α and β parameters of the GaLoM to be less than or equal to an upper bound. An application of the same method to the error parameters of the BLIM (Doignon & Falmagne, 1999; Falmagne & Doignon, 1988a) is described in Stefanutti and Robusto (2009). The MATLAB code for estimating and testing the CoGaLoM is provided in Appendix A.3.

3.1.1 The log-barrier method

The log-barrier method is one of the interior point methods, which are a class of algorithms for solving linear and non-linear convex optimization problems which reach an optimal solution by traversing the interior of a feasible region. To illustrate the log-barrier method, an example is chosen which introduces the application in the CoGaLoM.

Consider the problem of minimizing a non-linear function $f(y)$ subject to the inequality constraints $y_q \geq 0$ and $y_q \leq y^*$. By applying the log-barrier method, the function to be minimized takes on the form:

$$h(y, \mu) = f(y) - \mu \sum_{q=1}^{n} \ln[y_q(y^* - y_q)],$$

where $\mu \geq 0$ is a penalization parameter. Note that h is convex on the feasible region.

In each iteration of the EM algorithm, the penalization parameter μ is gradually decreased by some amount (say, $\mu_{t+1} = c\mu$, for $0 < c < 1$). It can be seen that, as μ tends to 0, the local minimizer of $h(y, \mu)$ approaches that of $f(y)$. It can also be seen that, as y_q approaches one of the boundaries of the feasible region (i.e., 0 or y^*), $y_q(y^* - y_q)$ tends to 0 and $h(y, \mu)$ approaches $+\infty$, thus providing a "barrier" to crossing the boundary.

Consider a situation in which the initial values of the parameter estimates belong to the feasible region (i.e., all the inequality constraints are satisfied at the outset). If the local minimizer of f is an interior point of the feasible region, then the barrier algorithm will reach such a point. If the local minimizer of f lies outside the feasible region, then a point belonging to the boundary of the region will be reached.

The aforementioned concepts are illustrated with an example. Consider the quadratic function $f(y) = 2y^2 - 2y + 1$, whose minimizer is $y = .5$. The function $f(y)$ has to be minimized subject to the inequality constraints $y \geq 0$ and $y \leq y^*$. Two values for the upper bounds ($y^* = .25, .60$) and three values for the penalization parameter ($\mu = 1, .10, .01$) are considered. Figure 3.1 depicts the values of y (x axis) which minimize $h(y, \mu)$ for each value of the penalization parameter (the three curves in the diagrams). The left-hand diagram refers to the condition with upper bound .60, the right-hand diagram to the condition with upper bound .25. With $y^* = .60$, the minimizer of $f(y)$ is an interior point of the feasible region and, as μ decreases, the minimizer of $h(y, \mu)$ approaches that of $f(y)$ ($y = .33, .42, .48$ for $\mu = 1, .10, .01$, respectively). With $y^* = .25$, the minimizer of $f(y)$ lies outside the upper bound of the feasible region and, as μ decreases, the minimizer of $h(y, \mu)$ is a value closer to the upper bound ($y = .14, .19, .24$ for $\mu = 1, .10, .01$, respectively).

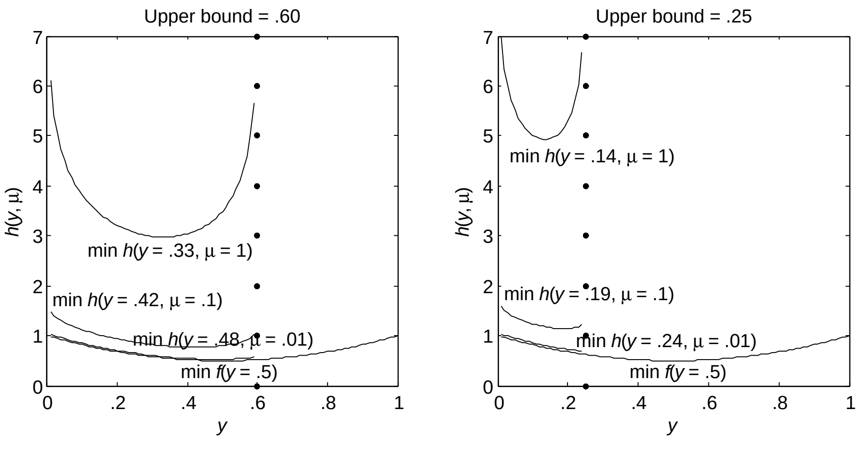

Figure 3.1: Minimization of the quadratic function $f(y)$ subject to the inequality constraints $y \geq 0$ and $y \leq y^*$. Two values for the upper bounds ($y^* = .25, .60$) and three values for the penalization parameter ($\mu = 1, .10, .01$) are considered.

3.1.2 Estimation of the careless error and lucky guess parameters in the Constrained Gain-Loss Model

In the CoGaLoM, the likelihood of the model is maximized subject to the constraint that the α and β parameters are less or equal to an upper bound. Only the part of the conditional expected log-likelihood $U(\boldsymbol{\theta}, \boldsymbol{\theta}')$ (Equation 2.3) that depends on the α_q and β_q parameters concretely undergoes a constrained maximization. This part is:

$$U_3(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{j=1}^{i} \sum_{C \subseteq S} \sum_{D \subseteq S} b_{jCD}[\ln P(\mathbf{x}_{1j}|C) + \ln P(\mathbf{x}_{2j}|D)].$$

It is worth recalling that $\boldsymbol{\theta}$ (resp. $\boldsymbol{\theta}'$) is the vector of model parameters obtained in the current (resp. previous) iteration of the EM algorithm, $\mathbf{x}_{1j}$ (resp. $\mathbf{x}_{2j}$) is a binary vector representing the response pattern of student j at time 1 (resp. time 2), C (resp. D) is a competence state at the pretest (resp. posttest), and b_{jCD} corresponds to $P(C, D|\mathbf{x}_{1j}, \mathbf{x}_{2j}, o_j, \boldsymbol{\theta}')$, where o_j is the learning object o student j has been presented with.

Given suitable upper bounds $\alpha_q^*, \beta_q^* \in [0, 1]$, there are four types of inequality constraints for each problem q: (a) $\alpha_q \geq 0$, (b) $\alpha_q \leq \alpha_q^*$, (c) $\beta_q \geq 0$, and (d) $\beta_q \leq \beta_q^*$, and the problem itself consists of maximizing the function $U_3(\boldsymbol{\theta}, \boldsymbol{\theta}')$ under the constraints (a) to (d). By an application of the log-barrier method, such a constrained maximization corresponds to an unconstrained minimization of the function:

$$V(\boldsymbol{\theta}, \boldsymbol{\theta}') = -\left\{ U_1(\boldsymbol{\theta}, \boldsymbol{\theta}') + \mu \sum_{q=1}^{n} \ln[\alpha_q(\alpha_q^* - \alpha_q)\beta_q(\beta_q^* - \beta_q)] \right\},$$

where μ is the penalization parameter introduced in the previous section.

In each iteration of the EM algorithm, the function $V(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is minimized by setting to zero its first partial derivatives with respect to the parameters α_q and β_q. The first partial derivative of $V(\boldsymbol{\theta}, \boldsymbol{\theta}')$ with respect to the parameter α_q turns out to be:

$$\frac{\partial V(\boldsymbol{\theta},\boldsymbol{\theta}')}{\partial \alpha_q} = \sum_{j=1}^{i}\sum_{C\subseteq S}\sum_{D\subseteq S} b_{jCD}\left[\frac{(1-x_{1jq})v_{qC}+(1-x_{2jq})v_{qD}}{\alpha_q} - \frac{x_{1jq}v_{qC}+x_{2jq}v_{qD}}{1-\alpha_q}\right] + \mu\frac{\alpha_q^*-2\alpha_q}{\alpha_q(\alpha_q^*-\alpha_q)}, \quad (3.1)$$

where $x_{1jq}\in\{0,1\}$ (resp. x_{2jq}) is equal to 1 of student j solved problem q at the pretest (resp. posttest), and $v_{qC}\in\{0,1\}$ (resp. v_{qD}) is equal to 1 if problem q is solvable by state C (resp. D). By setting the right-hand term of (3.1) to zero, it follows that:

$$(1-\alpha_q)(\alpha_q^*-\alpha_q)\sum_{j=1}^{i}\sum_{C\subseteq S}\sum_{D\subseteq S} b_{jCD}[(1-x_{1jq})v_{qC}+(1-x_{2jq})v_{qD}]+\mu(1-\alpha_q)(\alpha_q^*-2\alpha_q) = \alpha_q(\alpha_q^*-\alpha_q)\sum_{j=1}^{i}\sum_{C\subseteq S}\sum_{D\subseteq S} b_{jCD}(x_{1jq}v_{qC}+x_{2jq}v_{qD}). \quad (3.2)$$

Solving for α_q a second degree equation is obtained. Let $a_{q,0} = \sum_{j=1}^{i}\sum_{C\subseteq S}\sum_{D\subseteq S} b_{jCD}[(1-x_{1jq})v_{qC}+(1-x_{2jq})v_{qD}]$, $a_{q,1} = \sum_{j=1}^{i}\sum_{C\subseteq S}\sum_{D\subseteq S} b_{jCD}(x_{1jq}v_{qC}+x_{2jq}v_{qD})$, and $a_q = a_{q,0}+a_{q,1}$. Between the two roots of Equation 3.2, the one which satisfies the constraint $0\le\alpha_q\le\alpha_q^*$ is:

$$\alpha_q = \frac{(a_{q,0}+2\mu)+\alpha_q^*(a_q+\mu)}{2(a_q+2\mu)} - \frac{\{[(a_{q,0}+2\mu)+\alpha_q^*(a_q+\mu)]^2-4(a_q+2\mu)\alpha_q^*(a_{q,0}+\mu)\}^{\frac{1}{2}}}{2(a_q+2\mu)}. \quad (3.3)$$

Equation 3.3 represents the adjustment of the estimates of parameter α_q in each iteration of the EM algorithm.

Following a similar development for the parameters β_q, and defining $b_{q,0} = \sum_{j=1}^{i}\sum_{C\subseteq S}\sum_{D\subseteq S} b_{jCD}[(1-x_{1jq})(1-v_{qC})+(1-x_{2jq})(1-v_{qD})]$, $b_{q,1} = \sum_{j=1}^{i}\sum_{C\subseteq S}\sum_{D\subseteq S} b_{jCD}[x_{1jq}(1-v_{qC})+x_{2jq}(1-v_{qD})]$, and $b_q = b_{q,0}+b_{q,1}$, the equation for adjustment of the estimates of such parameters under the constraint $0\le\beta_q\le\beta_q^*$ is:

$$\beta_q = \frac{(b_{q,1}+2\mu)+\beta_q^*(b_q+\mu)}{2(b_q+2\mu)} - \frac{\{[(b_{q,1}+2\mu)+\beta_q^*(b_q+\mu)]^2-4(b_q+2\mu)\beta_q^*(b_{q,1}+\mu)\}^{\frac{1}{2}}}{2(b_q+2\mu)}. \quad (3.4)$$

3.2 A simulation study

The primary objectives of the simulation study are to test the CoGaLoM with respect to the parameter identifiability and the recognizability of the skill assignment underlying the data. In particular, it investigates whether constraining the parameter space of the α and β estimates constitutes a way of: (a) reducing the variability of the whole set of parameters in non-identifiable models, and (b) facilitating the identification of the model which incorporates the skill assignment underlying the data. A secondary objective is to investigate the effects of misspecifications of the skill assignment on the recovery of the problem parameters and the identification of the skill profiles. Different levels of information and noise in the data were considered.

3.2.1 Simulation of data sets

Three thousand random data sets were generated according to the GaLoM. Six conditions were produced by considering two ratios between the number of problems and underlying skills, and three levels of noise in the data. Two collections with 10 and 20 problems were generated, and five skills were set to underlie both. Via the conjunctive model, each problem from the two collections was associated with the skills that were necessary and sufficient for its mastery according to the skill maps represented in Table 3.1. Each of the two resulting performance structures $\mathcal{K}_{c10}$ and $\mathcal{K}_{c20}$ contains 32 states, and was used to generate the data for the conditions with 10 and 20 problems, respectively. The noise in the data was set to be low (α_{true} and $\beta_{true} \leq .1$), medium (α_{true} and $\beta_{true} \leq .3$) or high (α_{true} and $\beta_{true} \leq .5$).

For each of the six conditions, 500 random data sets were simulated by using

Table 3.1: Skill Assignment in the Collections with 10 and 20 Problems

10 Problems collection		20 Problems collection			
Problem	Skills	Problem	Skills	Problem	Skills
1	$\{a\}$	1	$\{a\}$	11	$\{b, d\}$
2	$\{b\}$	2	$\{b\}$	12	$\{a, b, d\}$
3	$\{c\}$	3	$\{c\}$	13	$\{c, d, e\}$
4	$\{d\}$	4	$\{d\}$	14	$\{a, b, d, e\}$
5	$\{e\}$	5	$\{e\}$	15	$\{b, c, e\}$
6	$\{a, b\}$	6	$\{a, b\}$	16	$\{a, b, c\}$
7	$\{a, c\}$	7	$\{a, c\}$	17	$\{a, c, d\}$
8	$\{d, e\}$	8	$\{b, c\}$	18	$\{a, b, c, e\}$
9	$\{a, b, c\}$	9	$\{c, e\}$	19	$\{a, d, e\}$
10	$\{c, d, e\}$	10	$\{d, e\}$	20	$\{b, c, d, e\}$

Note. Letters from a to e refer to the five skills.

a parametric bootstrap (see, e.g., Langeheine et al., 1996; von Davier, 1997). Each data set takes into account the effects of four learning objects on the skills. The first learning object was set to poorly affect both gain and loss of the skills (γ_{true} and $\lambda_{true} \leq .33$), the second to highly affect gain and poorly affect loss ($\gamma_{true} \geq .66; \lambda_{true} \leq .33$), the third to poorly affect gain and highly affect loss ($\gamma_{true} \leq .33; \lambda_{true} \geq .66$), the fourth to highly affect both gain and loss (γ_{true} and $\lambda_{true} \geq .66$). The initial probabilities of the skills π_{true} were set to be in the interval $[.1, .9]$ because, as pointed out in the previous chapter, estimating gain and loss probabilities is both difficult and negligible when initial probabilities are very high or very low. The true model parameters were randomly generated according to the aforementioned constraints. The initial probabilities were fixed across all conditions, and the gain and loss parameters were fixed across the conditions concerning levels of noise and information. The number of response patterns was set to 1,000 for each data set (250 for each learning object).

3.2.2 Estimation of the models

For each of the 500×6 data sets, correct and incorrect models were estimated. The correct models incorporated the performance structures that were used to generate the data (i.e., $\mathcal{K}_{c10}$ and $\mathcal{K}_{c20}$). The incorrect models incorporated two performance structures that were different from the correct ones ($\mathcal{K}_{i10}$ and $\mathcal{K}_{i20}$). These performance structures were delineated by two skill maps obtained by modifying some of the skill assignments of the generating skill maps. The modifications are listed in Table 3.2. $\mathcal{K}_{i10}$ contains 32 performance states, $\mathcal{K}_{i20}$ contains 30 performance states.

Table 3.2: Misspecifications of the Skill Assignment in the Collections with 10 and 20 Problems

Problem	Original	Modified	N skills added	N skills deleted
		10 Problems collection		
6	$\{a, b\}$	$\{a, b, d\}$	1	0
8	$\{d, e\}$	$\{a, d, e\}$	1	0
9	$\{a, b, c\}$	$\{a, b\}$	0	1
10	$\{c, d, e\}$	$\{b, d, e\}$	1	1
		20 Problems collection		
4	$\{d\}$	$\{c, d\}$	1	0
7	$\{a, c\}$	$\{a, c, d, e\}$	2	0
9	$\{c, e\}$	$\{a, c, e\}$	1	0
13	$\{c, d, e\}$	$\{c, e\}$	0	1
14	$\{a, b, d, e\}$	$\{a, b, c, d, e\}$	1	0
18	$\{a, b, c, e\}$	$\{a, e\}$	0	2
20	$\{b, c, d, e\}$	$\{b, d, e\}$	0	1

Note. Letters from a to e refer to the five skills.

In order to have a measure of how much the incorrect models differed from the correct ones, a discrepancy index for performance structures was computed[1]. The discrepancy between $\mathcal{K}_{i10}$ and $\mathcal{K}_{c10}$ was .47 (SD = .71), and that between

[1]Doignon and Falmagne (1999) described the following measure of the distance between two performance structures.

Given two performance structures $\mathcal{K}$ and $\mathcal{K}'$ on the same domain Q, and two performance states $K \in \mathcal{K}$ and $K' \in \mathcal{K}'$, the symmetric difference between K and K' is defined as:

$$K \Delta K' = |(K \setminus K') \cup (K' \setminus K)|.$$

$K \Delta K'$ specifies the number of problems that are elements of either, but not both, sets K and K'. The distance of the state $K \in \mathcal{K}$ from the performance structure $\mathcal{K}'$ is then computed

$\mathcal{K}_{i20}$ and $\mathcal{K}_{c20}$ was .97 ($SD = .98$).

Both correct and incorrect models were estimated four times by the CoGaLoM, each time with a different choice of the upper bound of the α and β parameters ($\alpha^*, \beta^* = 1, .5, .3, .1$; the upper bound was equal for all problems). Note that, when $\alpha^*, \beta^* = 1$, the CoGaloM corresponds to the GaLoM.

3.2.3 Testing parameter identifiability, recognizability of the true skill assignment, and effects of misspecifications of the skills assignment

In total, 2 (10 and 20 problems) $\times$ 3 (noise $\leq .1, .3, .5$) $\times$ 2 (correct and incorrect models) $\times$ 4 ($\alpha^*, \beta^* = 1, .5, .3, .1$) conditions were produced.

Model identifiability was tested in the conditions in which the unconstrained models were estimated ($\alpha^*, \beta^* = 1$; GaLoM). For this purpose, one of the 500 simulated data sets was randomly selected and the model parameters were estimated 100 times, by randomly varying their initial values between 0 and 1. The α and β parameters were randomly generated between 0 and .5. An unconstrained model was taken to be identifiable when the standard deviations were less than .01 for all the parameters.

---

as:

$$d(K, \mathcal{K}') = \min\{K \Delta K' : K' \in \mathcal{K}'\}.$$

$d(K, \mathcal{K}')$ is the minimum of the symmetric differences between K and all the performance states $K' \in \mathcal{K}'$. The discrepancy index from $\mathcal{K}$ to $\mathcal{K}'$ is obtained by computing the mean of the minimum distances $d(K, \mathcal{K}')$ of the states $K \in \mathcal{K}$ from $\mathcal{K}'$:

$$D(\mathcal{K}, \mathcal{K}') = \frac{1}{|\mathcal{K}|} \sum_{K \in \mathcal{K}} d(K, \mathcal{K}').$$

$D(\mathcal{K}, \mathcal{K}')$ can be considered as a measure of how well a performance structure $\mathcal{K}$ approximates a performance structure $\mathcal{K}'$. It is clear that $D(\mathcal{K}, \mathcal{K})$ is equal to 0. It should also be observed that, in general, D is not commutative, that is, $D(\mathcal{K}, \mathcal{K}') \neq D(\mathcal{K}', \mathcal{K})$. Along with $D(\mathcal{K}, \mathcal{K}')$, a standard deviation $SD(\mathcal{K}, \mathcal{K}')$ can be computed.

The effect of constraining the α and β parameters on reducing the variability of the whole set of parameters has been investigated in the non-identifiable models. The procedure is similar to the one used for testing model identifiability with the difference that the α and β parameters were randomly generated between 0 and the upper bounds .5, .3, .1. A parameter was taken to be unique when its standard deviations were less than .01.

The recognizability of the true skill assignment was tested by computing the proportion $P(\chi_c^2 < \chi_i^2)$ of data sets in which the Pearson's Chi-square of the correct models was smaller than that of the incorrect models.

The effect of misspecifications of the skill assignment on the identification of the skill profiles was assessed in the conditions in which the unconstrained models were estimated. For each of the 1,000 observations (response patterns), the competence state was taken that was modal across the 500 simulations. The number of observations that were classified into each competence state by the correct and the incorrect models was computed. Moreover, an agreement rate was calculated for each competence state by taking the number of observations that the incorrect model classified into a competence state in accordance with the correct model, and dividing it by the number of observations that the correct model classified into that competence state. An agreement rate across the competence states was calculated as well. Thus, when all the classifications of the incorrect model are in agreement with those of the correct model, the agreement rates are 1. In the opposite case, they are 0.

The effect of misspecifications on the recovery of the problem parameters was tested by considering the bias between the true problem parameters and the mean of the problem estimates that were reproduced by the correct and incorrect models on the 500 simulated data sets.

3.2.4 Results

Concerning the problem of identifiability, the unconstrained ($\alpha^*, \beta^* = 1$; GaLoM) correct and incorrect models are identifiable (i.e., $SD < .01$ for all parameters) in both the conditions with 10 and 20 problems when noise is $\leq .1$, and they are identifiable in the condition with 20 problems when noise is $\leq .3$.

As expected, both the level of information and the noise in the data contribute to the non-identifiability of the models. The number of not uniquely determined ($SD \geq .01$) parameters increases with the level of noise in the data, and it is greater when the collection with 10 problems is considered. Interestingly, on equal levels of information and noise, this number is greater when the incorrect models are estimated. The smaller the level of the upper bound, the smaller the number of parameters with multiple solutions. With an upper bound of .1, the correct and incorrect models that were not identifiable in the unconstrained estimation reach a unique solution for the estimates of all their parameters. In the condition with 20 problems and noise $\leq .5$ correct model suffices an upper bound of .3 for reaching a unique solution for all the parameters. Therefore, constraining the parameter space of the α and β estimates only, reduces the variability of the whole set of model parameters.

Concerning the problem of recognizing the true skill assignment, Figure 3.2 depicts the proportion of simulated data sets in which the Chi-square of the correct model χ^2_c was smaller than the Chi-square of the incorrect model χ^2_i (y axis) for each level of noise (x axis) and each value of upper bound (the four lines in the diagrams). The left-hand diagram refers to the conditions with 10 problems, the right-hand diagram to the conditions with 20 problems. With respect to the former, the proportion of cases in which the correct model obtains a better fit decreases when the noise increases from .1 to .3 and the upper bound is .3 or higher. This proportion keeps on decreasing when the noise reaches .5 and the upper bound is .5, whereas it unexpectedly increases when the upper bounds are 1 and .3. Regardless of the level of noise in the data, when the upper bound is .1 the correct model obtains a better fit in almost 100% of simulated data sets. In the conditions with 20 problems, regardless of the value of the upper bound, the higher the noise in the data, the lower the proportion of cases in which the correct model obtains a better fit. Moreover, regardless of the level of noise in the data, the smaller the value of the upper bound, the better the separation between the correct model and the incorrect one. Constraining the α and β estimates is therefore a way for increasing the recognizability of the true skill assignment.

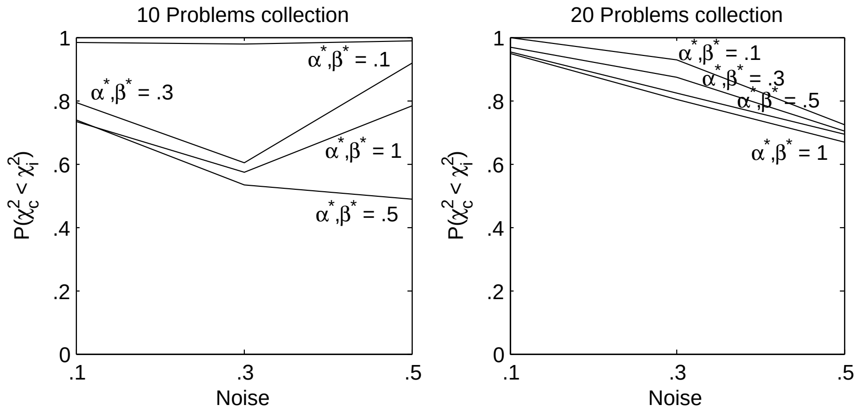

Figure 3.2: Proportion of simulated data sets in which the Chi-square of the correct model χ_c^2 turned out to be smaller than the Chi-square of the incorrect model χ_i^2 (y axis) for each level of noise (x axis) and each value of upper bound (the four lines in the diagrams).

The rest of this section concerns the effects of misspecifications of the skill assignment on the identification of the competence states and the recovery of the true problem parameters. Only the conditions in which the unconstrained models were identifiable are considered. Table 3.3 contains the number of observations classified into each competence state by the correct and the incorrect model in the condition with 20 problems and noise $\leq .1$, together with the agreement rates (AR). Note that the results concerning the correct model are not necessarily free of error. Nevertheless, they represent what is found under the optimal estimation condition.

A first thing to notice is that, for both the correct model and the incorrect model, the number of observations in each competence state at pretest reflects the initial probabilities of the skills and the number of observations at posttest reflects the effects of the four learning objects on the gain and loss probabilities. At pretest, the overall agreement rate between the two models is .90. Among the competence states with a non-small number of observations, the lower agreement between the two models is on state $\{b, c, d, e\}$ ($AR = .83$). This is probably due to the fact that this competence state represents a skill combination that is present in the generating skill map but not in the modified one (see Tables 3.1 and 3.2). At posttest, the two models are totally in agreement in classifying all the observations in the full competence state $\{a, b, c, d, e\}$ when the condition is $\gamma_{true} \geq .66$ and $\lambda_{true} \leq .33$. This result can be explained by considering that, for both the correct and the incorrect model, the full state is the most probable competence state at posttest given the initial probabilities of the skills and the learning object effect that is considered. The agreement rate between the two models is extremely low ($AR = .02$) in the condition $\gamma_{true} \leq .33$ and $\lambda_{true} \geq .66$.

The distribution of the observations in the condition with noise .3 is similar to that in the condition with noise .1 but the agreement between the correct and incorrect model is in general lower (see Table 3.4). The agreement between the correct and incorrect model become problematic in the condition with 10 problems and noise .1. In particular, the two models are totally in disagreement at pretest and in the condition γ_{true} and $\lambda_{true} \leq .33$ at posttest (see Table 3.5).

Figure 3.3 depicts the true problem parameters (x axis) plotted versus the mean of the problem parameter estimates (and the related standard errors) reproduced on the 500 simulated data sets (y axis) in the condition with 20 problems and noise .1. Left-hand diagrams refer to the correct model and right-hand diagrams to the incorrect model. Upper diagrams refer to the unconstrained estimation and lower diagrams to the estimation with upper bound of .1. When the estimation is unconstrained, the α and β probabilities of all problems are well-recovered in the correct model. In contrast, the α probabilities of problems 13, 18 and 20, and the β probabilities of problems 4, 7 and 9 are overestimated in the incorrect model. These biases are consistent with the modifications induced to the generating skill map (see Table 3.2). The overestimation of the α probability of problems 13, 18 and 20 is due to the deletion of some skills that were associated with these problems. The overestimation of the β probability of problems 4, 7 and 9 is due to the addition of some skills to these problems. Only problem 14 is not affected by the modification induced to its skill assignment. The recovery of problem parameters observed in the correct model with the upper bound of .1 does not differ from that observed with the unconstrained estimation. This is not the case with the incorrect model. Indeed, the estimates of the α and β parameters that were overestimated in the unconstrained estimation are now very close to the upper bound.

The recovery of problem parameters is similar in the condition with noise .3, but empirical biases of the estimates are larger for both the correct and incorrect models (see Figure 3.4). In the condition with 10 problems and noise .1, overestimations are observed which are consistent with the modifications induced to the generating skill map (see Figure 3.5 and Table 3.2). Moreover, overestimations are also observed for the α and β probabilities of some problems whose skill assignment was not modified. This result points to non-local effects of the misspecifications which are probably due to the fact that only a few of $2^{|S|}$ possible skill combinations could have been taken into account in a collection of 10 problems.

3.2.5 Discussion

The simulation study investigated whether constraining the parameter space of the α and β estimates was a way of reducing the variability of the whole set of

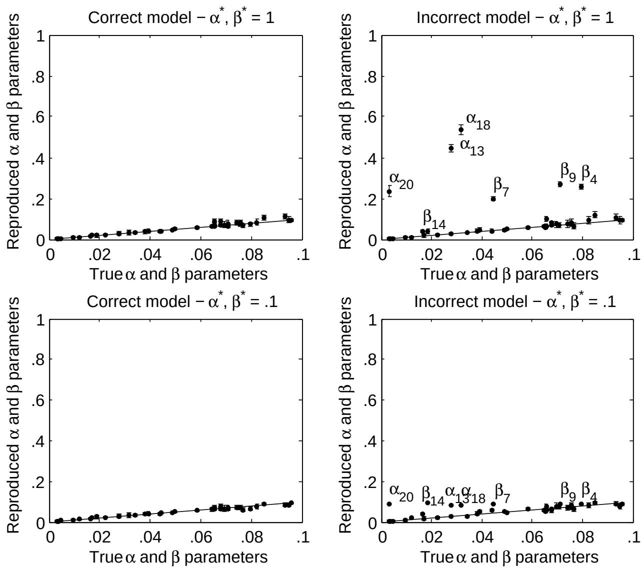

Figure 3.3: True problem parameters (x axis) versus mean of the problem parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) by both correct model (left-hand diagrams) and incorrect model (right-hand diagrams). Conditions with 20 problems, noise $\leq .1$, and upper bound $= 1$ (upper diagrams) and .1 (lower diagrams). The straight line $x = y$ is added for reference.

parameters in non-identifiable models, and of facilitating the identification of the skill assignment underlying the data. The effects of misspecifications of the skill assignment on the recovery of problem parameters and the identification of the skill profiles have been investigated as well.

As expected, model identifiability depends on both the level of information and the noise in the data. The simulations showed that, when the noise in the data is low (e.g., α_{true} and $\beta_{true} \leq .1$), the models might be identifiable even if the skills are assessed by means of a small number of problems. When the noise increases (e.g., α_{true} and $\beta_{true} \leq .3$), a greater number of problems is required for the models to be identifiable.

Across the different conditions, the number of non-identifiable parameters in-

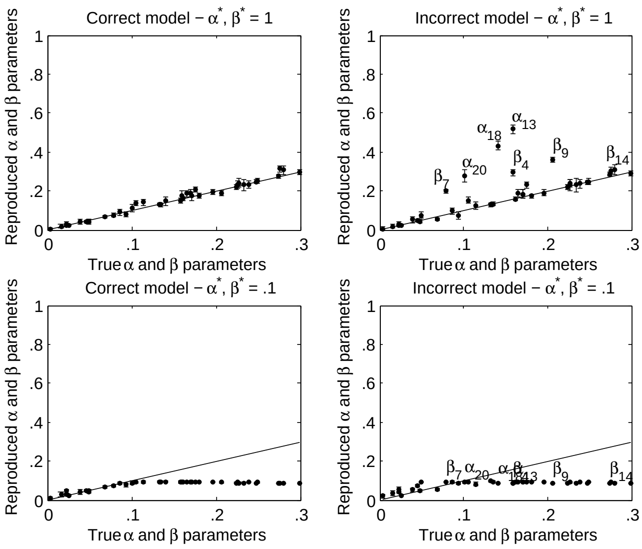

Figure 3.4: True problem parameters (x axis) versus mean of the problem parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) by both correct model (left-hand diagrams) and incorrect model (right-hand diagrams). Conditions with 20 problems, noise $\leq .3$, and upper bound $= 1$ (upper diagrams) and .1 (lower diagrams). The straight line $x = y$ is added for reference.

creased when the data became more noisy and the collection with 10 problems was considered. Interestingly, on equal levels of information and noise, this number was greater when the incorrect models were estimated. These results, together with those of the simulation study presented in the previous chapter, suggest that the data should provide enough information about the skills and that they should not be too noisy in order that the models are identifiable. Moreover, it is reasonable to expect that an important role in identifiability is played by the way the skills and the problems are related to each other.

Constraining the parameter space of the α and β estimates is a way of finding a unique solution for the entire set of the model parameters in non-identifiable models. The simulations showed that, as the upper bound got lower, the number

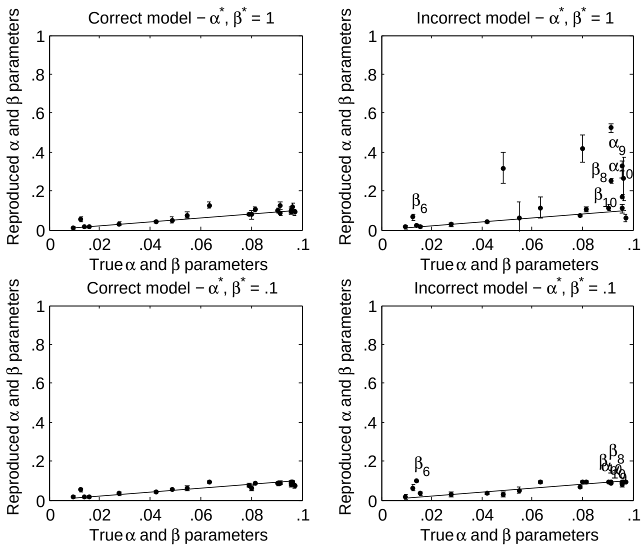

Figure 3.5: True problem parameters (x axis) versus mean of the problem parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) by both correct model (left-hand diagrams) and incorrect model (right-hand diagrams). Conditions with 10 problems, noise $\leq .1$, and upper bound $= 1$ (upper diagrams) and .1 (lower diagrams). The straight line $x = y$ is added for reference.

of non-identifiable parameters decreased until all parameters reached only one solution for their estimates. This is due to the fact that there is a connection between the α and β probabilities of the problems and the π, γ and λ probabilities of the skills that are associated with those problems. It should be noted that, through this procedure, biased estimates can be obtained. This is especially true for the α and β parameters which lie on the upper bound.

Misspecifications of the skill assignment affect the identification of the skill profiles and the recovery of the problem parameters. This result is in line with the findings of Rupp and Templin (2008), and it suggests the importance of properly associating the skills with the problems in order to reach an accurate assessment of students and the didactic interventions, and to obtain reliable estimates of the problems that are contained in the assessment instrument.

Constraining the parameter space of the α and β estimates also helps to recognize the skill assignment underlying the data. It is interesting to note that, regardless of the level of noise in the data, the best separation between the correct model and the incorrect one is obtained with the smallest upper bound value. This happens in spite of the fact that an upper bound lower than the true value of the α and β parameters will produce biased estimates for these parameters. This result can be understood by considering that, unlike the correct model, the incorrect one tends to inflate the α and β estimates of the problems whose skill assignment is incorrect until the model likelihood reaches its maximum value. By imposing an upper bound to such estimates, the CoGaLoM constrains the incorrect model more than the correct model. As a consequence, the likelihood of the incorrect model decreases faster than that of the correct one.

3.3 Final remarks

An extension of the GaLoM has been presented, which uses the log-barrier method in order to constrain the estimates of the careless error and lucky guess parameters to be less than or equal to an upper bound. However, other ways of constraining the parameters of a model have been proposed in the literature. Agresti and Lang (1993), Clogg (1979), Lazarsfeld and Henry (1968), and Lindsay, Clogg, and Grego (1991) proposed extensions of latent class models in which some of the conditional probabilities were fixed to given values or constrained to be equal. Using these kinds of constraints for the GaLoM might be difficult in practical applications in which there is neither prior knowledge about the error probability of a problem nor a valid reason for constraining different problems to have the same error probabilities. In this respect, the log-barrier method seems to be more feasible. Indeed, it imposes neither equality nor fixed-value constraints on the parameter estimates, but it only constrains them to be within a feasible region. Junker and Sijtsma (2001) proposed a monotonicity constraint for the careless error and lucky guess parameters of the DINA model (de la Torre & Douglas, 2004; Macready & Dayton, 1977). Given a problem q, the constraint requires that $\alpha_q + \beta_q < 1$. This approach does not prevent the inflation of error parameters that occurs when incorrect models are estimated, and it might not be useful in practical applications for recognizing the skills assignment that underlies the data.

In the simulation study that has been presented, the correct models were tested and compared with the incorrect ones. This situation could not hold in practice. In fact, in practical applications the skill multimap underlying the data is not known. Moreover, when there is not much theory about which skills are measured by the assessment instrument and how they are related to the problems, more than one skill multimap could be plausible in theory, and no one corresponding to the true skill multimap. In these cases, the skill multimap that best approximates the true one has to be identified. The next chapter investigates the usefulness of constraining the estimates of the α and β parameters for recognizing, among a number of alternatives, the skill assignment that best approximates the true one.

A final remark concerns the complexity of the models that are compared. The complexity of a model is usually related to the number of parameters. Adding parameters to a model increases its complexity, and it may also improve its fit to the data. When the models to be compared have the same number of parameters, Pearson's Chi-square statistic can be used. This has been the case of the study presented. On the contrary, when the models to be compared have a different number of parameters, indices such as the AIC and the BIC should be used because they correct the fit for complexity.

However, there is a complication that concerns us when comparing different models. In the GaloM, model complexity depends on the number of problems, skills and learning objects. Even so, it is not excluded that models incorporating performance structures with a different numbers of states result in different levels of complexity even if the overall number of parameters is the same. Given a set of skills S, there are $2^{|S|}$ possible competence states. If the function between the competence states and the performance states that are delineated from them is one-to-one, different competence states will delineate different performance states. The performance structure will therefore contain $2^{|S|}$ states. If the function is not one-to-one, different competence states will delineate the same performance state, and the performance structure will contain less than $2^{|S|}$. As an example, consider the correct and incorrect models concerning the condition with 20 problems in the study presented. The two models have the same number of parameters (55 parameters). The correct model incorporated a structure containing 32 performance states. It means that, through the generating skill map, each of the $2^5 = 32$ competence states delineated a different performance state. The incorrect model incorporated a structure containing 30 performance states. Through the modified skill map, both the competence states $\{\}$ and $\{d\}$ delineated the performance state in which no problem is solved, and both the competence states $\{a\}$ and $\{a, d\}$ delineated the performance state in which only problem 1 is solved. A greater number of performance states might lead to a model that fits better than another even if the two models have the same number of parameters. This form of complexity should be taken into account when comparing different models.

Table 3.3: Classification of the Observations within the Competence States and Agreement Rates for the Unconstrained Correct and Incorrect Models in the Condition with 20 Problems and Noise $\leq .1$

Competence states	Correct model	Incorrect model	AR
	Pretest		
$\{b, c\}$	104	151	.90
$\{b, e\}$	1	1	1.00
$\{b, c, d\}$	4	7	.50
$\{b, c, e\}$	867	799	.91
$\{a, b, c, e\}$	1	2	.00
$\{b, c, d, e\}$	23	40	.83
			.90
	Posttest - γ_{true} and $\lambda_{true} \leq .33$		
$\{b, c\}$	1	0	.00
$\{c, e\}$	5	0	.00
$\{b, c, e\}$	71	51	.69
$\{c, d, e\}$	3	11	1.00
$\{b, c, d, e\}$	170	188	.99
			.88
	Posttest - $\gamma_{true} \geq .66; \lambda_{true} \leq .33$		
$\{a, b, c, d, e\}$	250	250	1.00
			1.00
	Posttest - $\gamma_{true} \leq .33; \lambda_{true} \geq .66$		
$\{\}$	250	5	.02
$\{d\}$	0	245	
			.02
	Posttest - γ_{true} and $\lambda_{true} \geq .66$		
$\{a\}$	6	2	.17
$\{a, d\}$	6	57	.67
$\{a, e\}$	114	52	.45
$\{a, d, e\}$	124	139	.85
			.64

Note. Letters from a to e refer to the five skills. AR = agreement rate.

Table 3.4: Classification of the Observations within the Competence States and Agreement Rates for the Unconstrained Correct and Incorrect Models in the Condition with 20 Problems and Noise $\leq .3$

Competence states	Correct model	Incorrect model	AR
	Pretest		
$\{b\}$	2	0	.00
$\{b, c\}$	78	186	.69
$\{c, e\}$	1	3	1.00
$\{b, c, d\}$	0	8	
$\{b, c, e\}$	890	600	.65
$\{c, d, e\}$	0	1	
$\{a, b, c, e\}$	1	2	.00
$\{b, c, d, e\}$	28	200	.71
			.66
	Posttest - γ_{true} and $\lambda_{true} \leq .33$		
$\{b, c\}$	2	13	1.00
$\{c, e\}$	9	1	.11
$\{b, c, d\}$	0	3	
$\{b, c, e\}$	56	19	.29
$\{c, d, e\}$	2	2	.50
$\{b, c, d, e\}$	181	212	.97
			.78
	Posttest - $\gamma_{true} \geq .66; \lambda_{true} \leq .33$		
$\{a, b, c, d, e\}$	250	250	1.00
			1.00
	Posttest - $\gamma_{true} \leq .33; \lambda_{true} \geq .66$		
$\{\}$	250	107	.43
$\{d\}$	0	135	
$\{d, e\}$		8	
			.43
	Posttest - γ_{true} and $\lambda_{true} \geq .66$		
$\{a\}$	6	2	.33
$\{a, d\}$	13	70	.92
$\{a, e\}$	48	16	.29
$\{a, d, e\}$	183	162	.80
			.70

Note. Letters from a to e refer to the five skills. AR = agreement rate.

Table 3.5: Classification of the Observations within the Competence States and Agreement Rates for the Unconstrained Correct and Incorrect Models in the Condition with 10 Problems and Noise $\leq .1$

Competence states	Correct model	Incorrect model	AR
	Pretest		
$\{b, c\}$	34	0	.00
$\{c, e\}$	3	0	.00
$\{b, c, d\}$	0	100	
$\{b, c, e\}$	957	0	.00
$\{b, c, d, e\}$	6	0	.00
			.00
	Posttest - γ_{true} and $\lambda_{true} \leq .33$		
$\{c, d\}$	0	23	
$\{c, e\}$	6	0	.00
$\{b, c, d\}$	0	226	
$\{b, c, e\}$	101	0	.00
$\{c, d, e\}$	13	0	.00
$\{a, b, c, d\}$	0	1	
$\{b, c, d, e\}$	130	0	.00
			.00
	Posttest - $\gamma_{true} \geq .66; \lambda_{true} \leq .33$		
$\{a, b, c, d, e\}$	250	250	1.00
			1.00
	Posttest - $\gamma_{true} \leq .33; \lambda_{true} \geq .66$		
$\{\}$	250	250	1.00
			1.00
	Posttest - γ_{true} and $\lambda_{true} \geq .66$		
$\{a\}$	5	169	1.00
$\{a, d\}$	3	0	.00
$\{a, e\}$	106	24	.19
$\{a, d, e\}$	136	57	.38
			.30

Note. Letters from a to e refer to the five skills. AR = agreement rate.

Chapter 4

Selection of the best skill assignment in practice with the Constrained Gain-Loss Model

The skill multimap has to be specified in an appropriate way for the assessment to be accurate and trustworthy. In practical applications, the skill multimap underlying the data is not known. Moreover, when there is not much theory about which skills are measured by the assessment instrument and how they are related to the problems, more than one skill multimap could be plausible, and the problem would be how to select the best one.

Standard statistics, such as the Pearson's Chi-square or the likelihood ratio can be used for computing the fit of models incorporating different skill multimaps. The skill multimap is then chosen which is associated to the model with the best fit. This approach does not guarantee that the best model is identified. As observed by Stefanutti and Robusto (2009), an incorrectly specified model can obtain a good fit, even better than that of the correct model, by an ad hoc inflation of the error probabilities of the problems. Large values of the error probabilities might be a sign of misfit that should be taken into account when comparing different models.

A possibility in this direction is to choose the model with fewer problems with large α or β probabilities, given that its model fit is acceptable. Since misspecifications of the skill multimap with respect to some problems inflate the estimate of their error probabilities, it is expected that the fewer the misspecifications in the skill multimap, the fewer the problems with large error probabilities. This approach, although appealing, might be not feasible in practice. In practical applications, the real level of noise in the data is unknown. Some devices can be used for reducing the occurrence of careless errors and lucky guesses in responding to the problems. For example, the problems might be designed to be open response, and there could be no time pressure imposed in responding to the problems. However, there could always be some unpredictable or difficult to control elements which affect the responses to some problems. These might concern, for example, the wording of a problem, the fatigue, and the motivational decrease. If this is the case, high error probabilities for some problems could be erroneously interpreted as a sign of misspecification of the skill multimap with respect to those problems.

In the previous chapter, an approach was proposed which compares the fit to the data of models which undergo a constrained estimation of their error probabilities. An important feature of this approach is that it takes into account the information about model fit which derives from both the standard fit statistics and the estimates of the error probabilities. The rationale behind the approach is the following. Since an incorrect model, unlike the correct model, tends to inflate the error estimates of the problems until the model likelihood reaches its maximum value, it is constrained more than the correct model when an upper bound is imposed on the error estimates. As a consequence, the fit of the incorrect model deteriorates to a greater extent than that of the correct model.

The usefulness of the approach has been demonstrated in distinguishing the correct model from an incorrect one. In this chapter a situation is described in which the correct model does not belong to the collection of models which are compared to each other. This is a more interesting situation to test the approach on because in practical applications it might be that none of the skill multimaps at hand corresponds to the true skill multimap underlying the data, and that the skill multimap that best approximates the true skill multimap has to be identified.

The chapter is organized as follows. In the next paragraph, a simulation study is presented which tests the possibility of identifying, among a number of incorrect models, the one that best approximates the correct model. Then, an empirical application is presented in which models that are entirely derived from the analysis of the problems and from precise assumptions about the skills required for solving them are compared with models in which some skill assignments are random. Finally, a procedure is described that in practical applications can help to interpret high error probabilities of the problems as a sign of noise in the data rather than misspecifications of the skill multimap.

4.1 A simulation study

The simulation study investigates whether constraining the estimates of the α and β parameters allows to identify, from a number of alternative models, the model that best approximates that underlying the data. In particular, given a collection of models at different distances from the one used for generating the data, whether the CoGaLoM can be used for recognizing the closest one is investigated. The recovery of the problem parameters and the identification of the skill profiles are investigated as well.

Different strategies for associating the problems with the skills, and levels of noise in the data were considered.

4.1.1 Simulation of data sets

Three thousand random data sets were generated according to the GaLoM. Six conditions were produced by considering two strategies for associating the problems with the skills, and three levels of noise in the data. A collection with 20 problems was considered, and five skills were set to underlie it. The skill map and the skill multimap that are represented in Table 4.1 were used to associate the problems with the skills via the conjunctive and the competency model, respectively. The two resulting performance structures ($\mathcal{K}_c$ denotes the one delineated by the skill map, $\mathcal{M}_c$ the one delineated by the skill multimap) contain 32 performance states, and they were used to generate the data for the conditions concerning the conjunctive and the competency model. The noise in the data was set to be low (α_{true} and $\beta_{true} \leq .1$), medium (α_{true} and $\beta_{true} \leq .3$) or high (α_{true} and $\beta_{true} \leq .5$).

For each of the six conditions, 500 random data sets were simulated by using a parametric bootstrap (see, e.g., Langeheine et al., 1996; von Davier, 1997). The number of response patterns was set to 1,000 for each data set. The initial probabilities of the skills π_{true} were set to be in the interval $[.1, .9]$, and the learning object was set to highly affect gain of the skills and poorly affect loss ($\gamma_{true} \geq .66; \lambda_{true} \leq .33$). The true model parameters were randomly generated according to the aforementioned constraints. To facilitate the comparisons among the conditions, the initial probabilities, and the gain and loss probabilities were fixed across all the conditions. The careless error and lucky guess probabilities were fixed across the conditions concerning the conjunctive and the competency model.

4.1.2 Estimation of the models

For each of the 500×6 data sets, correct and incorrect models were estimated. The correct models (henceforth denoted by K_c and M_c) incorporated the performance

Table 4.1: Skill Assignment in the Conjunctive and Competency Model in the Simulation Study

Conjunctive model		Competency model	
Problem	Competencies	Problem	Competencies
1	$\{a\}$	1	$\{a\}$
2	$\{b\}$	2	$\{b\}$
3	$\{c\}$	3	$\{c\}$
4	$\{d\}$	4	$\{d\}$
5	$\{e\}$	5	$\{e\}$
6	$\{a, b\}$	6	$\{a, b\}, \{b, d, e\}$
7	$\{a, c\}$	7	$\{a, c\}$
8	$\{b, c\}$	8	$\{b, c\}$
9	$\{c, e\}$	9	$\{c, e\}, \{b, d, e\}$
10	$\{d, e\}$	10	$\{d, e\}, \{a, d\}$
11	$\{b, d\}$	11	$\{b, d\}$
12	$\{a, b, d\}$	12	$\{a, b, d\}$
13	$\{c, d, e\}$	13	$\{c, d, e\}$
14	$\{a, b, d, e\}$	14	$\{a, b, d, e\}, \{a, c, e\}$
15	$\{b, c, e\}$	15	$\{b, c, e\}$
16	$\{a, b, c\}$	16	$\{a, b, c\}, \{b, c, d, e\}$
17	$\{a, c, d\}$	17	$\{a, c, d\}$
18	$\{a, b, c, e\}$	18	$\{a, b, c, e\}$
19	$\{a, d, e\}$	19	$\{a, d, e\}$
20	$\{b, c, d, e\}$	20	$\{b, c, d, e\}$

Note. Letters from a to e refer to the five skills.

structures that were used to generate the data ($\mathcal{K}_c$ and $\mathcal{M}_c$). Incorrect models were created that incorporated performance structures at increasing distance from the correct ones. These performance structures were obtained in the following way. With respect to the conjunctive models, three skill maps were created by modifying the generating one (see Table 4.2). In a skill map, only the skills assigned to a problem were modified. Besides this modification, in another skill map the skills assigned to a second problem were also modified. Besides the previous modifications, in the last skill map, the skills assigned to a third problem were also modified. Therefore, the three skill maps differed from the generating skill map for the skills assigned to one, two or three problems. In a similar way, three skill multimaps were created that differed from the generating skill multimap for the skills assigned to one, two or three problems (see Table 4.2). The performance structures delineated by the three skill maps ($\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$) and those delineated by the three skill multimaps ($\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$) contain 32 performance states as the correct performance structures (the subscript indicates the number of problems whose skill assignment was modified).

The incorrect performance structures were at increasing distance from correct ones. The discrepancies between $\mathcal{K}_c$ and $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$ were respectively .25, .38, and .50 ($SD = .43, .60, .75$, respectively). The discrepancies between $\mathcal{M}_c$ and $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ were respectively .19, .25, and .38 ($SD = .39, .43, .65$, respectively). The incorrect conjunctive models are denoted by K_1, K_2 and K_3. The incorrect competency models are denoted by M_1, M_2 and M_3.

Both correct and incorrect models were estimated three times by the CoGaLoM, each time with a different choice of the upper bound of the α and β parameters ($\alpha^*, \beta^* = 1, .5, .1$; the upper bound was equal for all problems). It worth remembering that, when $\alpha^*, \beta^* = 1$, the CoGaloM corresponds to the GaLoM.

Table 4.2: Misspecifications of the Skill Assignments in the Conjunctive and Competency Models in the Simulation Study

Problem	Models	Original	Modified	N skills added	N skills deleted
		Conjunctive models			
6	$\mathrm{K}_1, \mathrm{K}_2, \mathrm{K}_3$	$\{a, b\}$	$\{a, d\}$	1	1
10	$\mathrm{K}_2, \mathrm{K}_3$	$\{d, e\}$	$\{b, d, e\}$	1	0
16	K_3	$\{a, b, c\}$	$\{a, c, e\}$	1	1
		Competency models			
10	$\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3$	$\{a, d\}$	$\{a, c, e\}$	2	1
14	M_3	$\{a, c, e\}$	$\{a, b, c\}$	1	1
16	$\mathrm{M}_2, \mathrm{M}_3$	$\{b, c, d, e\}$	$\{a, c, d, e\}$	1	1

Note. Letters from a to e refer to the five skills. $\mathrm{K}_1, \mathrm{K}_2$ and K_3 (resp. $\mathrm{M}_1, \mathrm{M}_2$ and M_3) = modified conjunctive (resp. competency) models with changes in the skills assigned to 1, 2 and 3 problems, respectively.

4.1.3 Testing model identifiability, effects of misspecifications of the skill assignment, and recognizability of the best skill assignment

Model identifiability was tested in the conditions in which the unconstrained models were estimated ($\alpha^*, \beta^* = 1$; GaLoM). For this purpose, one of the 500 simulated data sets was randomly selected and the model parameters were estimated 100 times, by randomly varying their initial values between 0 and 1. The α and β parameters were randomly generated between 0 and .5. A model was taken to be identifiable when the standard deviations were less than .01 for all the parameters.

The effect of misspecifications of the skills assignment on the identification of the skill profiles and the recovery of the problem parameters was tested in the conditions in which the unconstrained models were estimated. The former was assessed by computing the number of observations that the correct and incorrect models classified into each competence state. Moreover, agreement rates were computed between the correct and incorrect models. Agreement rates were calculated for each competence state by taking the number of observations that the incorrect models classified into a competence state in accordance with the correct models, and dividing it by the number of observations that the correct models classified into that competence state. Agreement rates across the competence states were also calculated. When all the classifications of the incorrect models were in agreement with those of the correct models, the agreement rates are 1. In the opposite case, they are 0.

The effect of misspecifications on the recovery of the problem parameters was tested by considering the bias between the true problems parameters and the mean of the problem parameter estimates reproduced by the correct and incorrect models on the 500 simulated data sets.

The recognizability of the skill assignment that best approximates the true one was assessed in all the conditions by computing the proportion of data sets in which the Pearson's Chi-square of the incorrect models at smaller distance from the correct ones was smaller than that of the incorrect models at greater distance. An average Chi-square was also computed for the incorrect models.

4.1.4 Results

All the unconstrained models are identifiable (i.e., $SD < .01$ for all parameters) in every noise condition, with the exception of the incorrect competency model M_3 in the condition with noise $\leq .5$ ($SD \geq .01$ for five out of 55 parameters). Among the competency models considered in this study, M_3 is the farthest from the correct one (M_c). The model reaches a unique solution for all the parameters with an upper bound of .1.

With respect to the effect of misspecifications of the skill assignment on the identification of the skill profiles, Table 4.3 contains the number of observations that the unconstrained conjunctive models classified into the competence states and the agreement rates (AR) between them. Table 4.4 contains the same information for the competency models. Note that the results concerning the correct models are not necessarily free of error (misclassifications are not excluded, especially when the noise is high). Nevertheless, they represent what is found under the optimal estimation condition. Results concerning the competency model M_3 in the condition with noise .5 are not provided because the model was not identifiable. Only results concerning the pretest are presented. In the posttest, all the conjunctive and competency models classified the entire set of observations in the full competence state $\{a, b, c, d, e\}$ ($AR_{c1} = AR_{c2} = AR_{c3} = 1$), regardless of the level of noise in the data. This result is consistent with the learning object effect that is considered in the study.

In general, there are no big differences in terms of how the conjunctive models categorize the observations in the pretest. The same thing goes for the compe-

Table 4.3: Classification of the Observations within the Competence States and Agreement Rates for the Unconstrained Conjunctive Models in the Pretest

Competence states	K_c	K_1	K_2	K_3	AR_{c1}	AR_{c2}	AR_{c3}
		Noise $\leq .1$					
$\{b, c\}$	128	160	159	91	1.00	.98	.69
$\{b, c, d\}$	13	12	11	15	.85	.77	.85
$\{b, c, e\}$	840	809	811	873	.96	.96	.99
$\{a, b, c, e\}$	1	1	1	1	1.00	1.00	1.00
$\{b, c, d, e\}$	18	18	18	20	.94	.94	1.00
					.97	.96	.95
		Noise $\leq .3$					
$\{b, c\}$	30	31	27	4	.73	.70	.13
$\{b, c, d\}$	0	1	1	0			
$\{b, c, e\}$	970	967	970	995	.99	.99	1.00
$\{a, b, c, e\}$	0	0	1	1			
$\{b, c, d, e\}$	0	1	1	0			
					.98	.98	.97
		Noise $\leq .5$					
$\{a, b, c\}$	1	3	3	4	1.00	1.00	1.00
$\{b, c, d\}$	512	566	534	355	.91	.88	.60
$\{b, c, e\}$	29	29	23	4	.59	.52	.14
$\{a, b, c, d\}$	451	391	431	636	.78	.83	.91
$\{a, b, c, e\}$	1	3	2	0	1.00	1.00	.00
$\{b, c, d, e\}$	6	8	7	1	.50	.50	.17
					.84	.84	.73

Note. Letters from a to e refer to the five skills. K_c = generating model; K_1, K_2 and K_3 = modified models with changes in the skills assigned to 1, 2 and 3 problems, respectively; AR_{c1}, AR_{c2} and AR_{c3} = agreement rate between K_c and K_1, K_2 and K_3, respectively.
In the posttest, the models classified all the observations in the competence state $\{a, b, c, d, e\}$, and $AR_{c1} = AR_{c2} = AR_{c3} = 1$.

tency models. This is probably due to the fact that the modifications induced to the generating skill assignments only concerned the skills associated with a few problems and, above all, they did not eliminate skill combinations corresponding to competence states observed in the simulated data sets. As expected, the more the incorrect models are far from the correct ones, the more the agreement rates decrease. This is more evident when the noise increases.

The true problem parameters (x axis) are plotted versus the mean of the problem parameter estimates (and the related standard errors) reproduced on the

Table 4.4: Classification of the Observations within the Competence States and Agreement Rates for the Unconstrained Competency Models in the Pretest

Competence states	M_c	M_1	M_2	M_3	AR_{c1}	AR_{c2}	AR_{c3}
		Noise $\leq .1$					
$\{b, c\}$	124	124	128	131	.98	.98	.97
$\{b, c, d\}$	16	15	14	16	.94	.88	.94
$\{b, c, e\}$	841	842	839	834	1.00	.99	.99
$\{a, b, c, e\}$	1	1	1	1	1.00	1.00	1.00
$\{b, c, d, e\}$	18	18	18	18	1.00	1.00	1.00
					.99	.99	.98
		Noise $\leq .3$					
$\{b, c\}$	42	72	75	79	.93	.95	.90
$\{b, c, d\}$	2	2	4	2	1.00	1.00	.00
$\{b, c, e\}$	956	926	921	918	.97	.96	.96
$\{a, b, c, e\}$	0	0	0	1			
					.96	.96	.95
		Noise $\leq .5$					
$\{a, b, c\}$	4	0	0		.00	.00	
$\{b, c, d\}$	968	968	931		.98	.94	
$\{b, c, e\}$	24	6	9		.21	.17	
$\{a, b, c, d\}$	4	26	60		.25	.75	
					.95	.92	

Note. Letters from a to e refer to the five skills. M_c = generating model; M_1, M_2 and M_3 = modified models with changes in the skills assigned to 1, 2 and 3 problems, respectively; AR_{c1}, AR_{c2} and AR_{c3} = agreement rate between M_c and M_1, M_2 and M_3, respectively.
Model M_3 was not identifiable in the condition with noise $\leq .5$. In the posttest, the models classified all the observations in the competence state $\{a, b, c, d, e\}$, and $AR_{c1} = AR_{c2} = AR_{c3} = 1$.

500 simulated data sets (y axis) by the unconstrained correct and incorrect models in all noise conditions. Figures 4.1 and 4.2 depict the results concerning the conjunctive models and the competency models, respectively. Results concerning the competency model M_3 in the condition with noise .5 are not provided because the model was not identifiable.

In the correct models K_c and M_c, the α and β parameters of all problems are well-recovered when the noise is .1 and .3, whereas those of only a few problems are overestimated when the noise is .5. In the incorrect models, overestimations of α and β estimates are observed which are consistent with all misspecifications

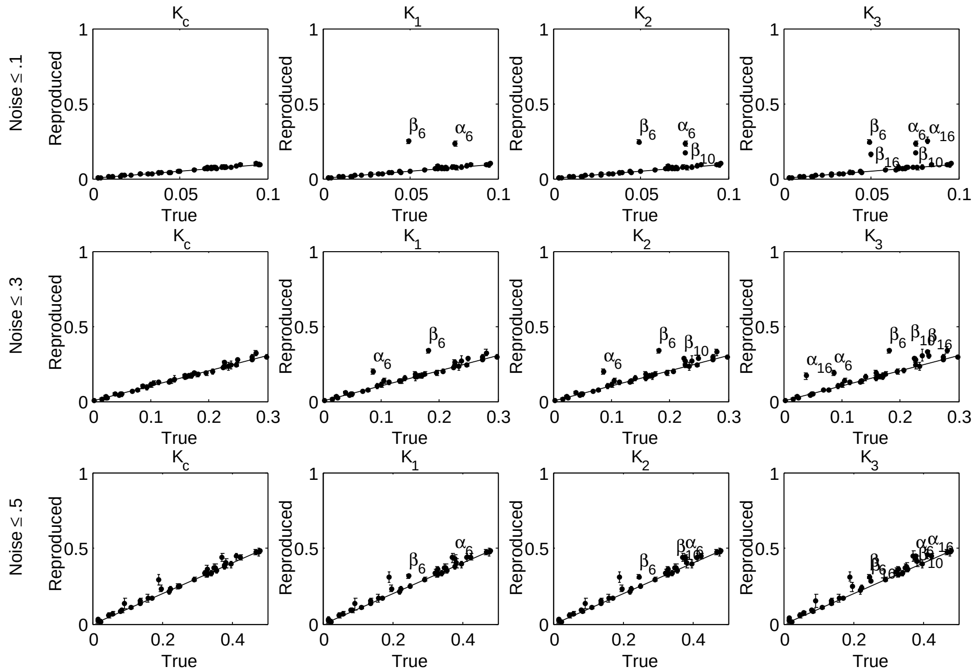

Figure 4.1: True problem parameters (x axis) versus mean of the problem parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) by the unconstrained correct (K_c) and incorrect (K_1, K_2, K_3) conjunctive models. The straight line $x = y$ is added for reference.

induced to the generating skill assignments (see Table 4.2). These overestimations are present in all levels of noise, but their size decreases with the increasing of the noise.

When the noise in the data is low, the incorrect model that has fewer problems with large α and β probabilities can be easily recognized, among alternative models, as the one that best approximates the correct model. In the condition with noise .1, K_1 is easily recognized, among the conjunctive models, as the one that best approximates K_c. Similarly, M_1 is easily recognized, among the competency models, as the one that best approximates M_c. When the noise in the data is high, it might be difficult to distinguish high values of α and β parameters which express misspecifications of the skill assignment from those

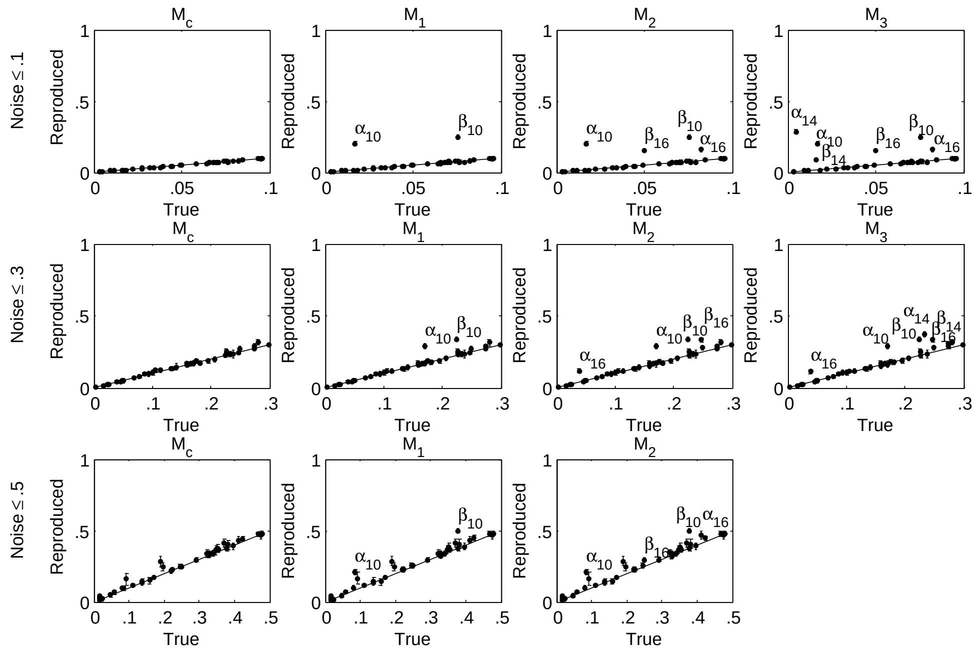

Figure 4.2: True problem parameters (x axis) versus mean of the problem parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) by the unconstrained correct (M_c) and incorrect (M_1, M_2, M_3) competency models. The straight line $x = y$ is added for reference.

which express noise.

This part concerns the possibility of recognizing the incorrect model that best approximates the correct one by imposing an upper bound on to the estimates of the α and β parameters. Figure 4.3 compares the fit of the incorrect conjunctive models for each level of noise (x axis) and each value of the upper bound (the three lines in the diagrams). Figure 4.4 contains the same information for the incorrect competency models. In both the figures, the upper diagrams depict the proportion of simulated data sets in which the Chi-square of a model at a smaller distance from the correct one was smaller than that of a model at greater distance. The lower diagrams depict the logarithm of the ratio between the average Chi-square of a model at greater distance from the correct one and that of a model at smaller distance. A value greater that 0 indicates that, between the two models that are compared, the one that is closest to the correct model is correctly identified as the best model. A value smaller than 0 indicates that the model that is farthest from the correct model is incorrectly identified as the best model.

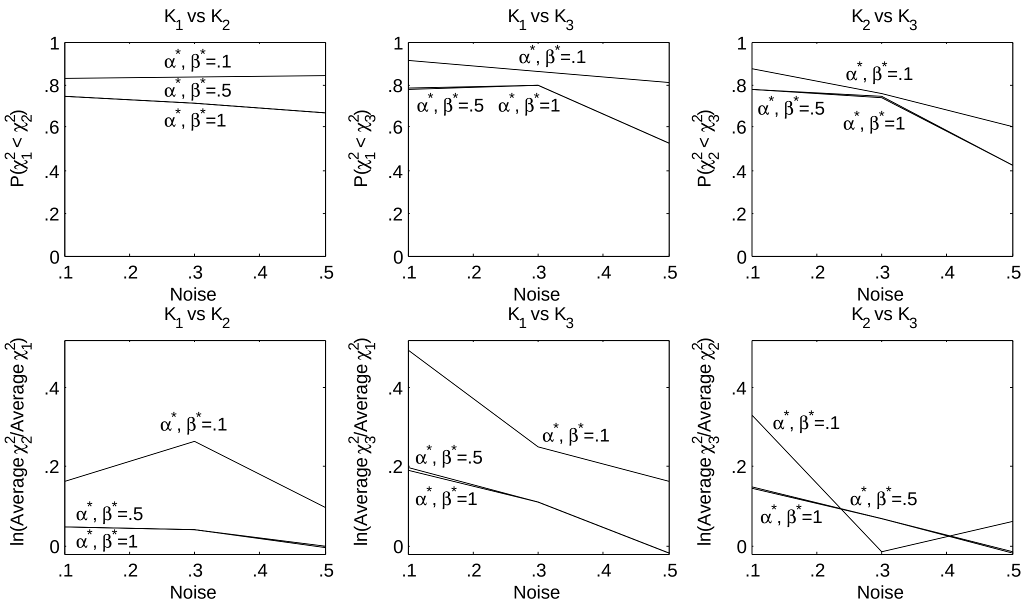

Figure 4.3: Comparison between the incorrect conjunctive models for each level of noise (x axis) and each value of the upper bound (the three lines in the diagrams). The upper diagrams depict the proportion of simulated data sets in which the Chi-square of a model that is closest to the correct one turned out to be smaller than that of a model that is farthest. The lower diagrams depict the logarithm of the ratio between the average Chi-square of a model that is farthest from the correct model and that of a model that is closest.

Some comments on the diagrams follow. In general, the recognizability of the best model decreases with the increase of the noise in the data. Usually, the best model is more easily recognized when the upper bound is set to .1. When the noise is .5 and the estimation of the α and β parameters is unconstrained, a model that is closest to the correct one might be not distinguished from a model that is farthest from it by simply comparing their fit. In these cases, an upper bound of

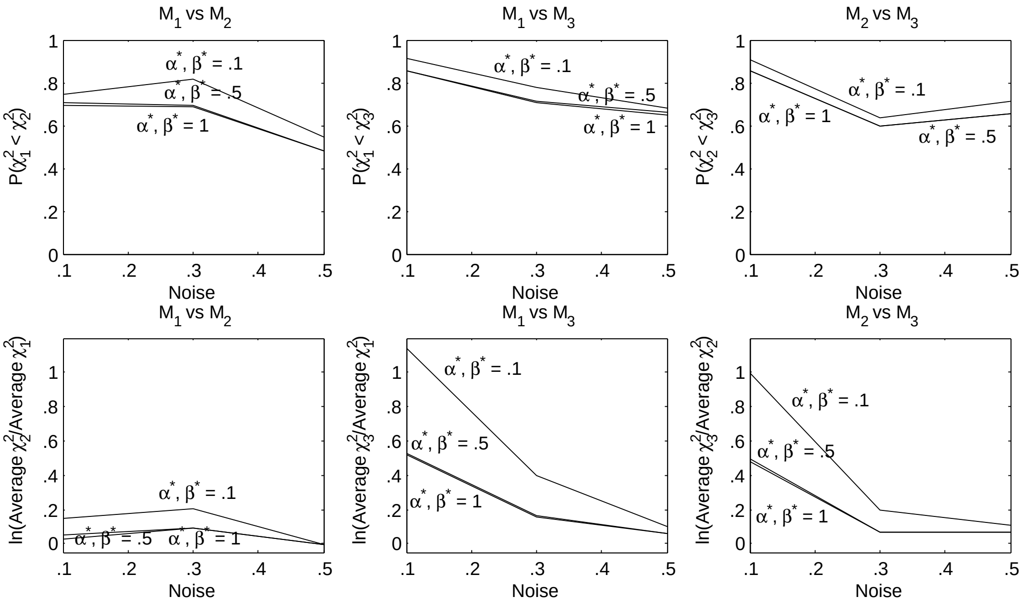

Figure 4.4: Comparison between the incorrect competency models for each level of noise (x axis) and each value of the upper bound (the three lines in the diagrams). The upper diagrams depict the proportion of simulated data sets in which the Chi-square of a model that is closest to the correct one turned out to be smaller than that of a model that is farthest. The lower diagrams depict the logarithm of the ratio between the average Chi-square of a model that is farthest from the correct model and that of a model that is closest.

.1 might affect the fit of the two models in such a way that the one that is closest to the correct one can be correctly identified. Finally, the results obtained with an upper bound of .5 do not differ from those obtained with the unconstrained estimation. This is probably due to the fact that for none of the models that have been considered were the overestimations of the α and β parameters greater than .5. For this reason, the upper bound did not apply a real constraint on these models.

4.1.5 Discussion

The simulation study investigated whether constraining the α and β parameters was a way of identifying, among alternative models, the one that best approximates the true model. Three conjunctive models and three competency models were created that were at increasing distance from the conjunctive and competency models used for generating the data.

Misspecifications of the skill assignment affect the identification of the skill profiles. This result is in line with what was observed in Chapter 3. It is interesting to note that effects can also be observed when only the skill assignments of a few problems are misspecified.

Large values of the α and β parameters of some problems can point to misspecifications of the skill assignment with respect to these problems. The simulations showed that, when the noise in the data is low, the model that has fewer problems with large α and β probabilities can be easily recognized, among alternative models, as the one that best approximates the correct model. In contrast, when the noise is high, looking at the large α and β probabilities might be not a feasible way of recognizing the best model. In practical applications, both the level of noise and the skill assignment underlying the data are unknown. Therefore, it might not be possible to distinguish high values of the estimates of α and β parameters which express misspecifications of the model from those which express noise.

When the unconstrained models are estimated, comparing their fit does not ensure that the best model is identified. This is particularly true as the noise in the data increases. Constraining the estimates of the α and β parameters increases the recognizability of the model that best approximates the correct one. This approach seems to be quite powerful given that, in the simulations presented, it allowed models that differed only in the skills assigned to a few problems to be distinguished between. The simulations showed that, for an upper bound to be effective, it should be lower than the level of noise in the data. In general, regardless of the level of noise, the best separation between the models was obtained with the lowest value of the upper bound. In practical applications, the use of a low upper bound is advisable given that the level of noise in the data is unknown.

4.2 An empirical application

In practical applications the skill assignment underlying the data is unknown. Moreover, different solutions could be plausible for associating the problems with the skills, and the problem would be establishing which the best is.

The assumption is made that a model, which is completely derived from the analysis of the problems and from specific assumptions about the skills required for solving them, represents the association underlying the data better than a model in which some of the associations between problems and skills are random.

The study investigates whether constraining the α and β estimates helps to recognize models that are entirely derived from plausible assumptions about the skills required for solving the problems from models in which some skill assignments are random.

4.2.1 Method

The study involved students at the University of Tübingen in the academic year 2009-2010. An invitation to participate in an internet-based study was sent to their email address. No financial reward was offered. The students who agreed to take part in the study were presented with two collections of 12 open response problems in elementary probability theory (see Appendix B.2). Four skills (determination of the probability of an event, probability of the complement of an event, stochastic independence, union of mutually exclusive events), and their combinations, were assumed to be required for solving the problems. The problems of the two collections were equivalent with respect to the content and the difficulty of the computation required.

After responding to the first collection of problems (pretest), the students were presented with a learning object consisting of concepts relative to the four skills and application examples (see Appendix B.2). Then, the students were presented with the second collection of problems (posttest).

From the students who took part in the study, 39 were selected that responded to all the problems and spent a certain time on the learning object pages and the test pages. Their mean age was 25.63 (SD = 7.51; range from 19 to 56), and 57 were female. Their responses to the problems were coded as correct (1) or incorrect (0).

4.2.2 Estimation of the models

Six distinct models were considered in the study. Two of them were completely derived from a systematic analysis of the content of the problems and from specific assumptions about the skills and the strategies that were required for solving them. They are therefore assumed to be two plausible models of the association between problems and skills. From each of these models, two models were derived by randomly modifying the skills assigned to certain problems. These four models are therefore assumed to be less plausible.

The first plausible model is a conjunctive model. It was delineated by a skill map associating each problem with the skills that were assumed to be necessary for solving it on the basis of what the problem requested (see Table 4.5). The performance structure (henceforth denoted by $\mathcal{K}_p$) delineated by the skill map contains 16 states. The second plausible model is a competency model. From the skill map, a skill multimap was derived by adding alternative solution strategies to some problems (see Table 4.5). These strategies were formulated by considering that the text of some problems provided the students with information that made it possible to use alternative approaches to solve the problem. The performance structure (henceforth denoted by $\mathcal{M}_p$) delineated by the skill multimap, contains 13 states. In the sequel, the plausible conjunctive and competency models are denoted by K_p and M_p, respectively.

Table 4.5: Skill Assignments in the Conjunctive and Competency Model in the Empirical Application

Conjunctive model		Competency model	
Problem	Competencies	Problem	Competencies
1	{pb}	1	{pb}
2	{cp}	2	{cp}, {un}
3	{un}	3	{un}, {cp}
4	{id}	4	{id}
5	{pb, cp}	5	{pb, cp}
6	{pb, cp}	6	{pb, cp}, {pb, un}
7	{pb, un}	7	{pb, un}, {pb, cp}
8	{pb, un}	8	{pb, un}
9	{pb, id}	9	{pb, id}
10	{cp, id}	10	{cp, id}
11	{pb, cp, id}	11	{pb, cp, id}
12	{pb, un, id}	12	{pb, un, id}, {pb, cp, id}

Note. cp = complement of an event; id = stochastic independence; pb = probability of an event; un = union of events.

From each of the two plausible models, two models were derived that were at increasing distance from them. With respect to the conjunctive models, two skill maps were created by randomly modifying the skills assigned to some problems in the plausible skill map (see Table 4.6). In the first one, only the skills assigned to a problem were modified. Besides this modification, in the second skill map the skills assigned to a second problem were also modified. Therefore, the two new skill maps differed from the plausible one due to the skills assigned to one or two problems. In a similar way, two skill multimaps were created that differed from the plausible skill multimap due to the skills assigned to one or two problems (see Table 4.6). The performance structures delineated by the two modified skill maps ($\mathcal{K}_1$, and $\mathcal{K}_2$) contain 16 performance states like $\mathcal{K}_p$, and the performance structures delineated by the two modified skill multimaps ($\mathcal{M}_1$, and $\mathcal{M}_2$) contain 13 performance states like $\mathcal{M}_p$ (the subscript indicates the number of problems for which skill assignment was modified).

Table 4.6: Misspecifications of the Skill Assignments in the Conjunctive and Competency Models in the Empirical Application

Problem	Models	Original	Modified	N skills added	N skills deleted
		Conjunctive models			
9	K_2	{pb, id}	{un, id}	1	1
11	$\mathrm{K}_1, \mathrm{K}_2$	{pb, cp, id}	{pb, cp, un}	1	1
		Competency models			
7	$\mathrm{M}_1, \mathrm{M}_2$	{pb, cp}	{pb, id}	1	1
12	M_2	{pb, cp, id}	{pb, cp, un}	1	1

Note. cp = complement of an event; id = stochastic independence; pb = probability of an event; un = union of events. K_1, and K_2 (resp. M_1 and M_2) = less plausible conjunctive (resp. competency) models with changes in the skills assigned to 1 and 2 problems, respectively.

The performance structures delineated by the modified skill maps and skill multimaps were at increasing distance from those delineated by the plausible skill map and skill multimap. The discrepancy between $\mathcal{K}_1$ and $\mathcal{K}_p$ was .13 (SD = .33) and that between $\mathcal{K}_2$ and $\mathcal{K}_p$ was .38 (SD = .60). The discrepancy between $\mathcal{M}_1$ and $\mathcal{M}_p$ was .15 (SD = .36) and that between $\mathcal{M}_2$ and $\mathcal{M}_p$ was .31 (SD = .46). In the sequel, the models incorporating those structures are denoted by K_1, K_2, M_1 and M_2.

The six models were estimated three times by the CoGaLoM, each time with a different choice of the upper bound of the α and β parameters ($\alpha^*, \beta^* = 1, .5, .1$; the upper bound was equal for all problems).

4.2.3 Testing model identifiability, goodness-of-fit, effects of modifications of the skill assignments, and recognizability of the plausible skill assignments

Identifiability and goodness-of-fit of the six models were tested in the conditions in which the estimation was unconstrained ($\alpha^*, \beta^* = 1$; GaLoM). To test model identifiability, parameters were estimated 100 times, by randomly varying their initial values between 0 and 1. The α and β parameters were randomly generated between 0 and .5. A model was taken to be identifiable when the standard deviations were less than .01 for all the parameters. Goodness-of-fit was tested using Pearson's Chi-square statistic and a parametric bootstrap (see Chapter 2 for details).

Also the effects of the modifications induced to the plausible skill assignments were tested when the estimation was unconstrained. The effects on the identification of the skill profiles were assessed by taking, for each student, the modal competence state in accordance with the estimated model. Moreover, agreement rates were computed between the plausible models and the less plausible ones. The effects of the modifications induced to the plausible skill assignments on the estimates of the problem parameters were assessed by analyzing whether the α and β probabilities obtained with the plausible models differed from those obtained with the modified models.

The recognizability of the plausible models was tested by comparing the Pearson's Chi-squares of plausible and less plausible models estimated with the different levels of upper bound. Moreover, 500 data sets were generated from the observed data sample by means of a non-parametric bootstrap. Their size was set to 39 response patterns, as for the observed data set. The six models were estimated on these data sets, and the proportion of data sets in which the Pearson's Chi-square of the plausible models was smaller than that of the less plausible models was computer. The proportion was also computed between the less plausible models at increasing distance from the plausible ones.

4.2.4 Results

All the unconstrained models are identifiable (i.e., $SD < .01$ for all parameters), and their goodness-of-fit is good (the proportion of random data sample whose Chi-square was less than the Chi-square of the observed data sample was .41, .49, .46, .47, .48 and .37 for the models K_p, K_1, K_2, M_p, M_1 and M_2, respectively).

Table 4.7 contains the unconstrained estimates of the parameters α and β obtained with the plausible models. In the conjunctive model K_p, the estimates of the careless error probabilities are rather high for problems 9 and 11 ($\alpha_9 = .40; \alpha_{11} = .39$), whereas those of the lucky guess probabilities are very high for problems 3, 7 and 9 ($\beta_3 = .59; \beta_7 = .79; \beta_9 = .40$). In the competency model M_p, additional solution strategies have been considered for the problems 2, 3, 6, 7 and 12. If they were really useful for solving the problems, a deflation of their lucky guess probability would be observed. The lucky guess probability of problem 3 decreases from .59 to $< .01$ with the addition of the solution strategy. This suggests that the strategy is really required for solving the problem. The lucky guess probability of problem 7 decreases from .79 to .50, suggesting that the solution strategy is somehow useful for solving the problem. Without considering problem 2, whose lucky guess probability was smaller than .01 in both the models, no improvements were observed for the lucky guess probabilities of the problems 6 and 12. It is difficult to establish whether this is because the alternative strategies were not useful for solving these problems, or because these probabilities are close to their actual lucky guess values. Interestingly, the solution strategies added to the skill map also affected the estimates of problem 8, whose skill assignment was not modified. Its lucky guess probability increased a bit and its careless error decreased. This non-local effect is due to the fact that, according to the skill map, problems 7 and 8 were clones of each other, that is, they were assigned to the same skills. The addition of a solution strategy to problem 7 could only have affected the estimates of the error probabilities of problem 8.

Given that the addition of the solution strategies decreased the lucky guess probabilities of some problems, the competency models can be assumed to approximate the true model underlying the data better than the conjunctive model.

Table 4.7: Maximum Likelihood Estimates of the α and β Parameters Obtained on the Unconstrained Plausible Conjunctive (K_p) and Competency (M_p) Models

	Careless error α		Lucky guess β	
Problem	K_p	M_p	K_p	M_p
1	.09	.09	.00	.00
2	.00	.00	.00	.00
3	.04	.05	.59	.00
4	.00	.00	.33	.33
5	.22	.21	.00	.00
6	.09	.09	.25	.25
7	.06	.05	.79	.50
8	.09	.00	.27	.36
9	.40	.40	.40	.40
10	.30	.31	.25	.25
11	.39	.39	.22	.22
12	.16	.21	.32	.31

Figure 4.5 depicts the problem parameter estimates obtained by the plausible conjunctive model plotted versus the problem estimates obtained by the less plausible conjunctive models. It was assumed that skills concerning *probability of an event*, *complement of an event* and *stochastic independence* were necessary and sufficient for solving problem 11 (see Table 4.5). The modifications induced to the plausible skill map replaced *stochastic independence* with *union of events* (see Table 4.6). If *stochastic independence* was necessary for solving problem 11, we would observe an inflation of its careless error probability in models K_1 and K_2. Moreover, if *union of events* was superfluous for solving problem 11, we would observe an inflation of its lucky guess probability. The results confirm our assumption that *stochastic independence* is necessary for solving the problem

($\alpha_{11} = .39, .53, .53$ in K_p, K_1, K_2, respectively). However, they suggest that the skill concerning *union of events* is not superfluous ($\beta_{11} = .22$ in K_p, $\beta_{11} \leq .01$ in K_1 and K_2). The modifications induced to problem 11 had an effect on the lucky guess estimates of problems 3, 7 and 8. It might be that the association with *union of events* somehow put problem 11 in connection with problems 3, 7 and 8, which were also associated with that skill. It was assumed that *probability of an event* and *stochastic independence* were necessary and sufficient to solve problem 9 (see Table 4.5). The modifications to the plausible skill map replaced *probability of an event* with *union of events* (see Table 4.6). In line with our assumptions, the results show that *probability of an event* is necessary for solving problem 9 ($\alpha_9 = .40, .44$ in K_p and K_3, respectively), and that *union of events* is not ($\beta_9 = .40, .46$ in K_p and K_3, respectively).

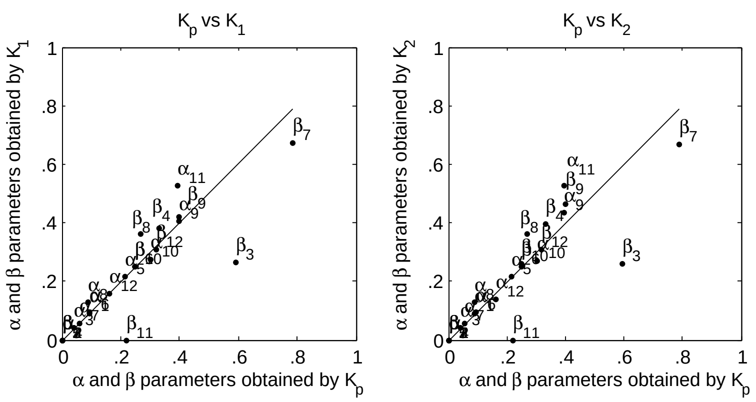

Figure 4.5: Problem parameter estimates obtained by the plausible (K_p) conjunctive model (x axis) versus problem parameter estimates obtained by the less plausible (K_1, K_2) conjunctive models (y axis) in the unconstrained estimation. The straight line $x = y$ is added for reference.

Figure 4.6 depicts the results concerning the competency models. It was assumed that two competencies were possible for solving problem 7, one containing *probability of an event* and *union of events*, and the other containing *probability of an event* and *complement of an event* (see Table 4.5). The latter competency was modified so that *complement of an event* was replaced with *stochastic independence* (see Table 4.6). If the competency was correctly specified, we would expect to observe an inflation of both careless error and lucky guess probabilities of problem 7 in models M_1 and M_2. The results showed an inflation of the lucky guess ($\beta_7 = .50, .65, .69$ in M_p, M_1, M_2, respectively), but not that of the careless error ($\alpha_7 = .05$ in all competency models). It was assumed that two competencies were possible for solving problem 12, one containing *probability of an event*, *union of events* and *stochastic independence*, and the other containing *probability of an event*, *complement of an event* and *stochastic independence* (see Table 4.5). The latter competency was modified so that *stochastic independence* was replaced with *union of events* (see Table 4.6). If the competency was correctly specified, we would expect to observe an inflation of both careless error and lucky guess probabilities of problem 12 in model M_2. Results confirmed our assumptions ($\alpha_{12} = .21, .38$ in M_p and M_2, respectively; $\beta_{12} = .31, .43$ in M_p and M_2, respectively). In general, the results suggest that the skill assignments of the plausible models are effectively more appropriate than those of the modified models.

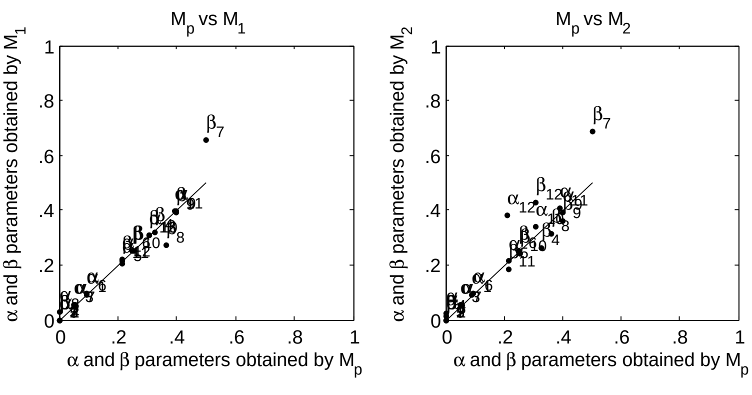

Figure 4.6: Problem parameter estimates obtained by the plausible (M_p) competency model (x axis) versus problem parameter estimates obtained by the less plausible (M_1, M_2) competency models (y axis) in the unconstrained estimation. The straight line $x = y$ is added for reference.

Table 4.8 contains the number of students classified into the competence states by the unconstrained conjunctive models, and the agreement rates (AR) between the models. Table 4.9 contains the same information for the competency models. Note that the results concerning the plausible models represent what is found when some plausible assumptions are made for associating the problems with the skills. Therefore, they are not necessarily correct.

Table 4.8: Classification of the Observations within the Competence States and Agreement Rates for the Unconstrained Conjunctive Models in the Pretest

Competence states	K_p	K_1	K_2	AR_{p1}	AR_{p2}
{pb}	1	1	1	1.00	1.00
{cp, un}	1	1	1	1.00	1.00
{pb, cp, un}	15	16	17	1.00	1.00
{pb, cp, id}	2	1	1	.50	.50
{pb, cp, un, id}	20	20	19	1.00	.95
				.97	.95

Note. cp = complement of an event; id = stochastic independence; pb = probability of an event; un = union of events. K_p = plausible model; K_1 and K_2 = less plausible models with changes in the skills assigned to 1 and 2 problems, respectively. AR_{p1}, AR_{p2} = agreement rate between the plausible and the less plausible models.
The results in the posttest are the same than those in the prestest.

Table 4.9: Classification of the Observations within the Competence States and Agreement Rates for the Unconstrained Competency Models in the Pretest

Competence states	M_p	M_1	M_2	AR_{p1}	AR_{p2}
{pb}	1	1	1	1.00	1.00
{pb, cp}	2	0	1	.00	.50
{cp, un}	1	1	1	1.00	1.00
{pb, cp, un}	12	14	12	1.00	1.00
{pb, cp, id}	5	5	6	1.00	1.00
{pb, cp, un, id}	18	18	18	1.00	1.00
				.95	.97

Note. cp = complement of an event; id = stochastic independence; pb = probability of an event; un = union of events. M_p = plausible model; M_1 and M_2 = less plausible models with changes in the skills assigned to 1 and 2 problems, respectively. AR_{p1}, AR_{p2} = agreement rate between the plausible and the less plausible models.
The results in the posttest are the same than those in the prestest.

The three conjunctive models do not differ in identifying the competence states of the students. The same thing goes for the three competency models. However, the small number of students who were considered in the analysis means that caution is required in interpreting these results.

This part concerns the possibility of distinguishing the plausible models from the less plausible ones by imposing an upper bound to the estimates of the α and β parameters. Figure 4.7 compares the fit of the three conjunctive models for each value of the upper bound (x axis). The upper diagram depicts the proportion of simulated data sets in which the Chi-square of K_p was smaller than that of K_1 and that of K_2. The proportion is also computed of data sets in which the Chi-square of K_1 was smaller that that of K_2. The lower diagram depicts the logarithm of the Chi-square of the models. Figure 4.8 contains the same information for the three competency models. When the estimation is unconstrained, the less plausible models can obtain a better fit than the plausible models. However, when an upper bound of .1 is specified, the plausible models fit the data better than the less plausible ones and, therefore, they can be correctly identified. The same thing goes when, between the less plausible models, that which is closer to the plausible one has to be identified.

The comparison between the Chi-square of the two plausible models highlights that, regardless of the level of the upper bound, M_p fits the data better than K_p. This is the case regardless of the fact that M_p incorporates a performance structure containing 3 states less than that incorporated by K_p. In line with what was observed with the estimates of the error probabilities, this result suggests that the plausible competency model M_p approximates the true model underlying the data better than the conjunctive model K_p.

4.2.5 Discussion

The study investigated whether constraining the α and β parameters allows the recognition of models that were entirely derived from precise assumptions about

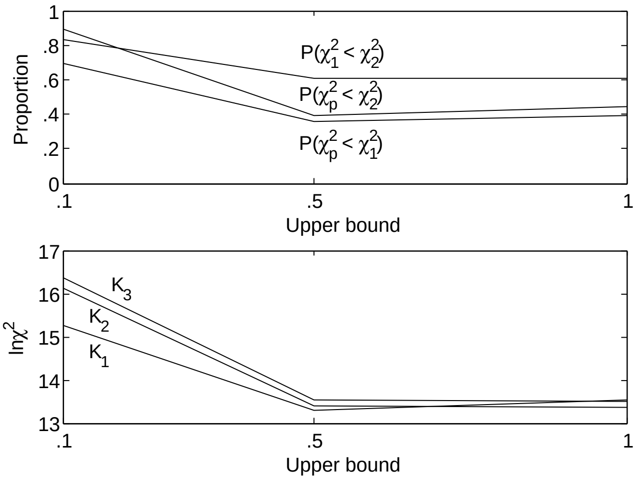

Figure 4.7: Comparison between the three conjunctive models for each value of the upper bound (x axis). The upper diagram depicts the proportion of simulated data sets in which the Chi-square of the plausible model was smaller than that of the less plausible models. The lower diagram depicts the logarithm of the Chi-square of the three models. K_p = plausible model; K_1 and K_2 = less plausible models with changes in the skills assigned to 1 and 2 problems, respectively.

the skills required for solving the problems from models in which some problems were randomly associated with the skills. The assumption was made that the former models represent the association between problems and skills which underlies the data better than the latter models.

The results show that constraining the estimates of the α and β parameters increases the probability that the plausible models are correctly identified when their fit is compared with that of less plausible models. This goes regardless of the fact that the models differed only in the skills assigned to a few problems.

A conjunctive model and a competency model were entirely derived from the analysis of the problems. The competency model resulted in a representation of the association between the problems and skills underlying the data that was better than the representation provided by the conjunctive model. However, there could be another model which describes the data even better. A possibility in this

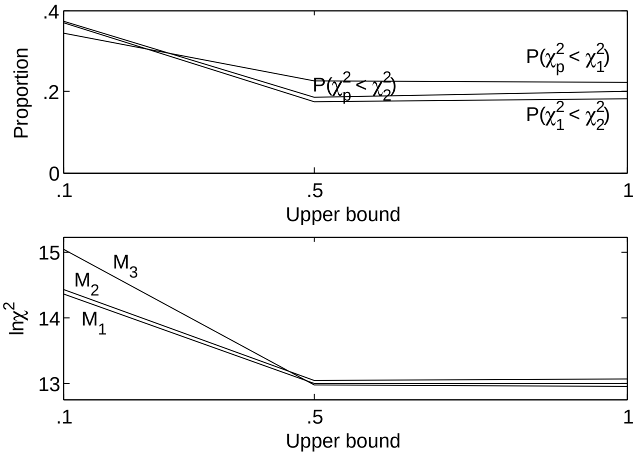

Figure 4.8: Comparison between the three competency models for each value of the upper bound (x axis). The upper diagram depicts the proportion of simulated data sets in which the Chi-square of the plausible model was smaller than that of the less plausible models. The lower diagram depicts the logarithm of the Chi-square of the three models. M_p = plausible model; M_1 and M_2 = less plausible models with changes in the skills assigned to 1 and 2 problems, respectively.

direction is the following. Problems 5 and 9 concern computations with a deck of cards. The careless error probability is quite high for both. It is reasonable to hypothesize that there are students who possess the skills required for solving the two problems, but that fail because they do not know what cards are in the deck. A skill could be added to these problems which concerns the knowledge of a deck of cards. This possibility has not been pursued here because it would lead to an increase in the number of parameters, and the number of problems would not be sufficient to obtain stable and trustworthy estimates.

4.3 Final remarks

An approach has been presented that allows the recognition of, among different models, the one that best describes the association between problems and skills which underlie the data. An important feature of this approach is that it takes into account the information about model fit which derives from both the standard fit statistics and the estimates of the error probabilities.

In practical applications, both the level of noise and the skill multimap underlying the data are unknown. Therefore, it is difficult to interpret large values of the careless error and the lucky guess probabilities as a sign of misspecification of the skill multimap rather than expression of noise. The following procedure can be used for this purpose. Having available a possibly large collection of skill multimaps, each one representing a way of associating problems with skills which is plausible in theory, the introduction of a reasonably small upper bound to the estimates of the error parameters, should allow the identification of the skill multimap that best approximates the true one. Once the best skill multimap has been identified, it is possible that the estimates of some α and β parameters of the model incorporating such skill multimap lie on the upper bound. If this is the case, these estimates could be biased. The model should therefore be estimated with no constraints. If there are large estimates of the α or β probabilities for some problems, these could be considered, with relative confidence, as a sign of noise in the data concerning those problems (e.g., bad wording of the problem) rather than failings in the specification of the skill multimap.

Chapter 5

A closer look at the learning process with a logistic reparametrization of the Gain-Loss Model

The GaLoM assesses learning processes by analyzing the effects of learning objects on specific skills. The gain and loss parameters are informative in this respect, because they specify the probabilities that the students attain and eventually lose specific skills as a result of the learning object they have been presented with. Given that the gain and loss parameters are indexed by learning object and skill, they inform the teachers of the effectiveness of the educational intervention in promoting specific skills, and enable them to select the best educational intervention for the specific needs of the students.

This chapter proposes a reparametrization of the GaLoM, which provides a new way of reading the information concerning the learning process. Logistic modelling is used to decompose both gain and loss parameters into parameters which describe the effects of the skills and the effects of the learning objects on the skills. Parameters that concern general effects of the learning objects are also derived. These new parameters provide information about the skills and the learning objects in a form that is easily accessible.

The chapter is organized in the following way. The next paragraph describes how the gain and loss parameters can be decomposed into a number of logistic parameters. Subsequently, the reparametrization is illustrated through an empirical application. The chapter concludes with some remarks concerning the proposed approach.

5.1 A logistic reparametrization of the Gain-Loss Model

Logistic modelling is used to decompose the gain and loss parameters into an effect of the skills and an effect of the learning objects on the skills. An overall effect of the learning objects is also derived. The MATLAB code for computing the logistic parameters is provided in Appendix A.4. The reader interested in logistic modelling is referred to Agresti (2002, 2007).

The probability γ_{os} that the students presented with learning object o gain skill s going from the pretest to the posttest is modelled as a logistic function of the difference between the effectiveness of learning object o in promoting the attainment of skill s, and the difficulty of gaining skill s. The logistic model takes on the form:

$$\ln \frac{\gamma_{os}}{1-\gamma_{os}} = \psi_{os} - \phi_s, \tag{5.1}$$

where $\ln \frac{\gamma_{os}}{1-\gamma_{os}}$ is the log-odds (or logit) of gaining skill s when learning object o is presented, ψ_{os} is the effectiveness of learning object o in promoting the attainment of skill s, and ϕ_s is the difficulty of gaining skill s. According to Equation 5.1, the more (resp. less) effective learning object o for the attainment of skill s, and the less (resp. more) difficult the gaining of skill s, the greater (resp. lesser) the probability of gaining skill s when learning object o is presented.

The parameters ϕ_s and ψ_{os} are defined as:

$$\phi_s = -\frac{\sum_{o=1}^{m} \ln[\gamma_{os}/(1-\gamma_{os})]}{m},$$

and

$$\psi_{os} = \ln[\gamma_{os}/(1-\gamma_{os})] + \phi_s,$$

where m is the number of learning objects.

The parameter ϕ_s is the negative of the average of the log-odds of skill s across the learning objects. It follows that a value greater (resp. smaller) than 0 means that, across the different learning objects that are considered, the probability of gaining skill s is smaller (resp. greater) than .5. Moreover, the greater the value of ϕ_s, the greater the difficulty of gaining skill s. Note that the difficulties of attaining different skills can be compared to each other without having to refer to a specific learning object. Parameter ψ_{os} is defined to represent a departure of the log-odds from ϕ_s. Since ϕ_s is an average across learning objects, it follows that $\sum_{o=1}^{m} \psi_{os} = 0$. The greater the value of ψ_{os}, the greater the effectiveness of learning object o in promoting the attainment of skill s. Note that the different difficulties of gaining the skills are taken into account when comparing the effects of a learning object on different skills.

A parameter ω_o representing the effectiveness of learning object o in promoting the attainment of knowledge in general can be computed as:

$$\omega_o = \frac{\sum_{s \in S} \psi_{os}}{|S|},$$

where $|S|$ is the cardinality of the set of skills.

The parameter ω_o is the average of parameters ψ_{os} across the skills. The greater the value of ω_o, the greater the overall effectiveness of learning object o in promoting the attainment of knowledge. Note that the effectiveness of different learning objects can be compared without having to refer to a specific skill.

The probability λ_{os} that the students presented with learning object o lose skill s going from the pretest to the posttest is modelled as a logistic function of the difference between the proclivity of losing skill s, and the effectiveness of learning object o in counteracting the loss of skill s. The model takes on the form:

$$\ln \frac{\lambda_{os}}{1-\lambda_{os}} = \phi'_s - \psi'_{os}, \tag{5.2}$$

where $\ln \frac{\lambda_{os}}{1-\lambda_{os}}$ is the log-odds of losing skill s when learning object o is presented, ϕ'_s is the proclivity of losing skill s, and ψ'_{os} is the effectiveness of learning object o in counteracting the loss of skill s. According to Equation 5.2, the lesser (resp. greater) the tendency of losing skill s, and the greater (resp. lesser) the effectiveness of learning object o in counteracting the loss of skill s, the smaller (resp. greater) the probability of losing skill s when learning object o is presented.

The parameters ϕ'_s and ψ'_{os} are defined as:

$$\phi'_s = \frac{\sum_{o=1}^{m} \ln[\lambda_{os}/(1-\lambda_{os})]}{m},$$

and

$$\psi'_{os} = -\ln[\lambda_{os}/(1-\lambda_{os})] + \phi'_s.$$

A value of ϕ'_s greater (resp. smaller) than 0 means that, across the different learning objects, the probability of losing skill s is greater (resp. smaller) than .5. However, the greater the value of ϕ'_s, the greater the proclivity of losing skill s. The greater the value of ψ'_{os}, the greater the effectiveness of learning object o in counteracting the loss of skill s. Note that for parameters ψ'_{os}, it also holds that $\sum_{o=1}^{m} \psi'_{os} = 0$.

It is worth specifying the meaning of parameters ϕ'_s and ψ'_{os} in practice. A high value for parameter ϕ'_s indicates that skill s is not yet consolidated in the students who already possess it. On the contrary, a high value for parameter ψ'_{os} indicates that learning object o is effective for consolidating skill s in the students who possess it.

A parameter ω'_o representing the effectiveness of learning object o in counteracting the loss of knowledge in general is computed as:

$$\omega'_o = \frac{\sum_{s \in S} \psi'_{os}}{|S|}.$$

The greater the value of ω'_o, the greater the overall effectiveness of learning object o in counteracting the loss of knowledge.

An example is now provided to illustrate how the logistic parameters should be interpreted. Five skills (denoted by letters from a to e) and three learning objects (LO_1, LO_2 and LO_3) are considered. The estimates of the gain and loss parameters which are presented are fictitious, and they were used for the sole purpose of illustrating particular situations which might be encountered in practice. Table 5.1 contains the estimates of the gamma parameters and the logistic parameters derived from them. The parameters ϕ are considered first. Skill d is the least difficult to be attained by the students who do not possess it in the pretest ($\phi_d = -1.55$), whereas skill e is the most difficult ($\phi_e = 1.11$). On the whole, learning object LO_1 is the most effective in promoting the attainment of the skills ($\omega_1 = .20$), whereas learning object LO_3 is the least effective ($\omega_3 = -.30$). LO_1 is more effective than LO_2 for the attainment of skills a, b and c, just as effective for the attainment of skill d, and less effective for the attainment of skill e. Differently, LO_1 is more effective than LO_3 for the attainment of all skills. A learning object can be considered to be at least adequate for the attainment of a skill if its effectiveness in promoting the attainment of that skill is greater than the difficulty of attaining it. LO_1 is adequate for the attainment of all skills, with the exception of skill e. Note that the probabilities of attaining skills a and b with LO_1 are the same. However, LO_1 is more effective on skill b than on skill a because attaining the former is more difficult.

Table 5.2 contains the estimates of the lambda parameters and the logistic parameters derived from them. All skills are consolidated in the students who possess them in the pretest. All learning objects are good in counteracting the

Table 5.1: Gain Parameters and Logistic Parameters Derived from Them

	Gain γ				ψ		
Skill	LO_1	LO_2	LO_3	ϕ	LO_1	LO_2	LO_3
a	.59	.56	.51	$-.22$	.15	.03	$-.18$
b	.59	.51	.49	$-.12$	.24	$-.08$	$-.16$
c	.55	.46	.39	.14	.34	$-.02$	$-.31$
d	.87	.87	.70	-1.55	.35	.35	$-.70$
e	.23	.30	.22	1.11	$-.10$	.26	$-.16$
				ω	.20	.11	$-.30$

Note. Letters from a to e refer to the five skills. LO_1, LO_2 and LO_3 refer to the three learning objects. ϕ = difficulty of gaining the skill; ω = effectiveness of the learning object in promoting the attainment of knowledge; ψ = effectiveness of the learning object in promoting the attainment of the skill.

loss of the skills, with learning object LO_1 being the best.

Table 5.2: Loss Parameters and Logistic Parameters Derived from Them

	Lossλ				ψ'		
Skill	LO_1	LO_2	LO_3	ϕ'	LO_1	LO_2	LO_3
a	.09	.11	.10	-2.20	.11	$-.11$	.00
b	.11	.12	.15	-1.94	.15	.05	$-.20$
c	.07	.06	.12	-2.44	.14	.31	$-.45$
d	.03	.03	.04	-3.38	.10	.10	$-.20$
e	.07	.08	.14	-2.28	.31	.16	$-.47$
				ω'	.16	.10	$-.26$

Note. Letters from a to e refer to the five skills. LO_1, LO_2 and LO_3 refer to the three learning objects. ϕ' = proclivity of losing a skill; ω' = effectiveness of a learning object in counteracting the loss of knowledge; ψ' = effectiveness of the learning object in counteracting the loss of a skill.

The example has shown what information concerning the learning process is derived from the logistic parameters. An important feature of these parameters is that their statistical significance can be tested. For this purpose, the parameters have to be standardized by dividing them by their standard errors. The values which are obtained are normally distributed. Moreover, the parameters can be compared to each other. A significance test of any difference between two parameters can be accomplished by dividing this difference by the square root of the sum of the squared standard errors of the two parameters.

5.2 An empirical application

The empirical application illustrates the proposed reparametrization. The purpose is to show the usefulness of the logistic parameters for highlighting features of the skills and the learning objects which are involved in the learning process.

5.2.1 Method

One hundred and seventy-two psychology students at the University of Padua participated in the study for course credits. The students were attending the course of Psychometrics in the academic year 2010-2011. Their mean age was 21.22 ($SD = 3.51$; range from 19 to 52), and 160 were female. The students were presented with two collections of 19 open response problems in elementary probability theory (see Appendix B.3). Four skills (determination of the probability of an event, probability of the complement of an event, stochastic independence, union of mutually exclusive events) were assumed to be required for solving the problems. The two collections (called forms A and B) were equivalent with respect to the content of the problems and the difficulty of the computations required.

A 2×2 experimental design with two learning objects (good and bad) and two assessment steps (pretest and posttest) was planned. The bad learning object only presented concepts of elementary probability theory concerning the four skills, whereas the good learning object also presented application examples (see Appendix B.3). In addition, both learning objects pointed out that the order in which the four skills were used was relevant for solving the problems. Since the examples provided by the good learning object were problems of the same nature as those administered to the students, this learning object was assumed to be more effective for solving the problems than the other one.

The data were collected in the classroom through a paper-and-pencil procedure. To reduce the occurrence of cheating, forms A and B were distributed in alternating order. After responding to the problems of the first collection (pretest), students belonging to a first group (Group G, $N = 90$) were presented with the good learning object, and those belonging to a second group (Group B, $N = 82$) with the bad learning object. Then, the students were presented with the second collection of problems (posttest). The students who received a form in the pretest (A or B) then received the other one in the posttest. The responses to the problems were coded as correct (1) or incorrect (0).

5.2.2 Model estimation

Different skill assignments were derived from the analysis of the problems and from specific assumptions about the skills that were required for solving them. The CoGaLoM and the procedure described in Chapter 4 were used in order to select the best skill assignment from the various alternatives. The skill assignment which was found to be the best is the skill multimap represented in Table 5.3.

Table 5.3: Skill Assignment in the Competency Model

Problem	Competencies	Problem	Competencies
1	{un}, {cp}	11	{pb, un}
2	{pb, cp}	12	{cp, id}
3	{pb, cp, id, so}	13	{pb, un}
4	{pb}	14	{cp}
5	{cp, id}	15	{pb, cp, un, id, so}
6	{id}	16	{pb, cp, id, so}
7	{pb, cp, un}	17	{pb, cp}, {pb, un}
8	{un, id}	18	{pb, un, id, so}
9	{pb, un, id, so}	19	{pb, cp}, {pb, un}
10	{pb, id, so}		

Note. cp = complement of an event; id = stochastic independence; pb = probability of an event; un = union of events; so = skill order.

In addition to the four skills concerning the concepts of elementary probability theory, a higher-order skill was considered which accounts for the capability of the students to identify the skills which are necessary for solving a problem and to use them in the proper order. This skill is called *correct order*. The order is not relevant for solving some problems. For example, consider problem 11 taken from form A: “Throw a dice. What is the probability of obtaining a 1 or a 4?” According to the specified skill multimap, the problem requires the skills concerning *probability of an event* and *union of events* to be solved. The problem can be solved by determining first the probability of the two events and then the probability of their union. Otherwise, it can be solved by determining first the successful event, and then its probability. The order is relevant for solving other problems. For example, consider problem 10 from form A: “A box contains 6 marbles of which 4 are white and 2 are black. Another box contains 8 marbles of which 3 are white and 5 are black. One marble is extracted randomly from each box. What is the probability of both marbles being white?” According to the skill multimap, the problem requires *probability of an event* and *stochastic independence* to be solved. To give the correct response, the probability of the two events has to be determined first. The skill *correct order* has been added to all problems that were associated with both *probability of an event* and *stochastic independence*, and it represents the knowledge of the students that the former skill has to be used first.

The performance structure delineated by the given skill multimap contains 19 performance states. The model incorporating this structure was estimated with no constraints (i.e., with the GaLoM).

5.2.3 Testing model identifiability, goodness-of-fit and significance of the logistic parameters

To test model identifiability, parameters were estimated 100 times, by randomly varying their initial values between 0 and 1. The error parameters were randomly generated between 0 and .5. The model was taken to be identifiable when the standard deviations were less than .01 for all parameters. Goodness-of-fit was tested using Pearson's Chi-square statistic and a parametric bootstrap with 500 replications (see Chapter 2 for details).

After estimating and validating the model, the logistic parameters were derived from the gain and loss parameters. The standard errors of the logistic parameters that were required to standardize and compare them were obtained by the parametric bootstrap.

5.2.4 Results

The model was identifiable (i.e., $SD < .01$ for all parameters), and its goodness-of-fit was good (the proportion of random data samples whose Chi-square was less than the Chi-square of the observed data sample was .55.

Table 5.4 contains the estimates of the parameters α and β. The careless error probabilities are quite small for all problems, with the exception of problem 15 ($\alpha_{15} = .37$). This result is not surprising if we consider that, according to the specified skill multimap, problem 15 requires the presence of all skills to be solved. The careless error probabilities of problems 9, 11, 15 and 16 increase if the skill *correct order* is eliminated, whereas their lucky guess probabilities do not change. This result shows that the skill is relevant for solving these problems. The lucky guess probabilities are not small for all problems, and they are very high for problems 2, 7, 11, 14 and 17 ($\beta_2 = .61; \beta_7 = .57; \beta_{11} = .43; \beta_{14} = .42; \beta_{17} = .44$). This might suggest failings in the specification of the skill multimap. However, the fact that the lucky guess probabilities are considerable for many problems suggests that the data are noisy due to the lucky guesses. This possibility is supported by the fact that the data were collected in a crowded classroom at the faculty, and this setting, together with the fact that course credits were given as a reward, might have encouraged the students to cheat in order to solve the greatest number of problems.

Table 5.4: Maximum Likelihood Estimates of the Parameters α and β

	Careless error		Lucky guess			Careless error		Lucky guess	
Problem	α	SE	β	SE	Problem	α	SE	β	SE
1	.09	.02	$<.01$	.25	11	.07	.02	.43	.06
2	.06	.02	.61	.06	12	.12	.03	.06	.03
3	.25	.04	.24	.04	13	.15	.03	.33	.05
4	.13	.02	.17	.08	14	.06	.02	.42	.13
5	.11	.02	.10	.04	15	.37	.05	.12	.02
6	.09	.02	.32	.05	16	.26	.04	.15	.03
7	.11	.02	.57	.05	17	.09	.02	.44	.07
8	.13	.03	.21	.04	18	.16	.04	.20	.03
9	.19	.04	.14	.03	19	.14	.02	.35	.08
10	.12	.03	.37	.04					

Note. Standard errors (SE) of the estimates were obtained by parametric bootstrap.

Table 5.5 contains the estimates of the parameters π, γ and λ. All skills have a high initial probability. This is consistent with the fact that the students had already been presented with the four concepts of elementary probability theory that are the subject of investigation, and they had already completed exercises on them before data collection took place. Of all of them, *stochastic independence* is the skill with the lowest initial probability ($\pi_{id} = .62$).

Table 5.5: Maximum Likelihood Estimates of the Parameters π, γ and λ

			Group G ($N = 90$)				Group B ($N = 82$)			
	Initial p.		Gain		Loss		Gain		Loss	
Skill	π	SE	γ	SE	λ	SE	γ	SE	λ	SE
Probability of an event	.86	.03	.39	.16	.02	.03	.22	.11	.06	.04
Complement of an event	.93	.02	.59	.31	.04	.05	.54	.31	.10	.07
Union of events	.81	.04	.49	.18	.12	.06	.07	.07	.03	.04
Stochastic independence	.62	.04	.61	.11	.01	.02	.34	.09	.01	.03
Skill order	.89	.04	.01	.18	$<.01$	.07	.21	.26	.12	.12

Note. Standard errors (SE) of the estimates were obtained by parametric bootstrap. Group G = good learning object; Group B = bad learning object.

The good learning object was more effective than the bad learning object in promoting the attainment of knowledge. The probabilities of gaining all four skills concerning elementary probability theory are greater in Group G than in Group B. This result confirms the assumption that providing application examples in addition to the basic concepts increases the probability of solving the problems. Unexpectedly, the probability of gaining the skill *correct order* is greater in Group B than in Group G.

There are small probabilities of losing the skills in both Groups G and B. This result might be due to compensation effects between the gain and loss parameters within each of the two groups. However, it might also be an expression of fatigue, decrease in motivation or boredom in the posttest for the students who solved the problems in the pretest.

The logistic parameters are now considered. All comparisons between the parameters that are described do not show statistically significant differences. However, they are presented for a descriptive purpose. Table 5.6 contains the logistic parameters derived from the gamma parameters, together with their standardized value. Among the skills concerning elementary probability theory, *union of events* is the most difficult to attain for the students who do not possess it in the pretest, whereas *complement of an event* is the least difficult. Among all skills, *correct order* is the most difficult to attain. This result is consistent with the nature of the skill. Given that it concerns the capability of the students to identify the skills which are necessary for solving a problem and to use them in the proper order, the skill requires exercise to be attained. On the whole, the good learning object was more effective than the bad learning object in promoting the attainment of the skills. This holds for all skills concerning elementary probability theory and, in particular, for *union of events*. The comparison between the effectiveness of the good learning object in promoting a skill and the difficulty of that skill shows that the learning object is adequate in promoting the *complement of an event* and *stochastic independence*. In contrast, the bad learning object is only adequate for promoting *complement of an event*. An interesting result can be found by comparing how effective the good learning object is on the different skills. Among the skills which were considered, the good learning object was more effective on *union of events*, even if it was not adequate for that skill, given its difficulty. This is due to the fact that attaining the skill is quite difficult for the students who did not possess it in the pretest.

Table 5.6: Logistic Parameters Derived from the Gain Parameters

			Group G ($N = 90$)		Group B ($N = 82$)	
Skill	ϕ	(*std*)	ψ	(*std*)	ψ	(*std*)
Probability of an event	.86	(.41)	.42	(.19)	−.42	(−.19)
Complement of an event	−.25	(−.05)	.11	(.02)	−.11	(−.02)
Union of events	1.35	(.47)	1.31	(.46)	−1.31	(−.46)
Stochastic independence	.11	(.21)	.57	(1.16)	−.57	(−1.16)
Correct order	3.21	(1.25)	−1.86	(−.66)	1.86	(.66)
	ω	(*std*)	.11	(.07)	−.11	(−.07)

Note. Group G = good learning object; Group B = bad learning object. ϕ = difficulty of gaining the skill; ω = effectiveness of the learning object in promoting the attainment of knowledge; ψ = effectiveness of the learning object in promoting the attainment of the skill.

Table 5.7 contains the logistic parameters derived from the lambda parameters, together with their standardized value. All skills are consolidated in the students who already possess them in the pretest. In particular, the probabilities of losing the skills concerning *complement of an event*, *stochastic independence*, and *correct order* are significantly smaller less than .5. On the whole, both learning objects are good for contrasting the loss of the skills, with the good learning object being the most effective one. The greatest difference between the two learning objects can be observed with respect to the skill concerning *correct order*.

Table 5.7: Logistic Parameters Derived from the Loss Parameters

			Group G ($N = 90$)		Group B ($N = 82$)	
Skill	ϕ'	(*std*)	ψ'	(*std*)	ψ'	(*std*)
Probability of an event	−3.23	(−1.70)	.48	(.25)	−.48	(−.25)
Complement of an event	**−2.67**	**(−2.23)**	.52	(.43)	−.52	(−.43)
Union of events	−2.74	(−1.01)	−.74	(−.27)	.74	(.27)
Stochastic independence	**−4.29**	**(−7.33)**	−.04	(−.07)	.04	(.07)
Correct order	**−3.94**	**(−2.00)**	1.95	(.91)	−1.95	(−.91)
	ω'	(*std*)	.43	(.51)	−.43	(−.51)

Note. The significant parameters are shown in bold. Group G = good learning object; Group B = bad learning object. ϕ' = proclivity of losing the skill; ω' = effectiveness of the learning object in counteracting the loss of knowledge; ψ' = effectiveness of the learning object in counteracting the loss of the skill.

5.2.5 Discussion

The empirical application illustrated the reparametrization of the gain and loss probabilities. In particular, it showed which information about the skills and the learning objects derives from the logistic parameters.

The logistic parameters derived from the gain probabilities allowed the identification of *union of event* and *correct order* as the most difficult skills to attain for the students who did not possess them, and to identify the good learning object as the most effective in promoting the attainment of knowledge. The analysis revealed the skills for which this learning object was adequate, and it provided information about its effectiveness on the different skills when their difficulty levels were taken into account. The logistic parameters derived from the loss probabilities demonstrated that the skills concerning *complement of an event*, *stochastic independence*, and *correct order* were very consolidated and stable in the students who possessed them, and that, in general, both learning objects were good at counteracting loss of the skills. The statistical significance of each parameter and each comparison between the parameters was tested.

5.3 Final remarks

A reparametrization of the GaLoM was presented, which allows a deeper analysis of the learning process. Some remarks about the proposed approach are now presented and discussed.

The GaLoM provides information about the learning process by means of $2m \times |S|$ gain and loss parameters, where m is the number of learning objects, and $|S|$ is the number of skills. The reparametrization allows $m(2 + 5 \times |S|)$ logistic parameters to be derived from the gain and loss parameters, which provide particular information about the skills and the learning objects involved in the learning process. It should be noted that this information is contained in the gain and loss parameters. However, deriving it through observation of these parameters might not be easy, especially when many learning objects and skills are taken into account. The reparametrization provides this information in a form that is easily accessible. This might be especially useful when evaluating and comparing a large number of skills and learning objects.

An important feature of the logistic parameters is that their statistical significance can be tested. Moreover, it is also possible to test the significance of the differences between them. In this chapter, the standard errors of the logistic parameters were obtained through a bootstrap procedure. Alternatively, they could be derived analytically. In testing the significance of the logistic parameters, the inflation of the type I error probability should be controlled for, and a unique value for the error probability should be attributed to all parameters of the same type. This value can be computed by dividing the nominal value of type I error probability (usually .05) by $2 \times z$, where 2 denotes a bidirectional hypothesis, and z denotes the number of linearly independent parameters. This number corresponds to $|S|$ for the parameters ϕ_s (i.e., all parameters ϕ are linearly independent), and to $|S| \times (m-1)$ for the parameters ψ_{os} (it is worth restating that $\sum_{o=1}^{m} \psi_{os} = 0$). The parameters ω_o are not independent, but they derive from ψ_{os}. Moreover, when many comparisons between the parameters are made, it is advisable to divide the type I error probability by the number of comparisons which are possible.

The proposed approach for the decomposition of the gain and loss probabilities can not be used when these probabilities are exactly 1 or 0. In the first case the denominator of the fraction in Equation 5.1 or 5.2 would be 0, whereas in the second case the logarithm of 0 should be taken. In both cases, an undefined value would be obtained. The analysis is possible by not considering the parameters concerning the same skill and the learning object of the parameter which is 0 or 1, and limiting it on the remaining parameters.

Chapter 6

Conclusions

The thesis has presented the GaLoM, which is a formal model for assessing learning processes. The model was developed in the context of formative assessment, an approach which has started to emerge in the last few decades. Formative assessment represents a sharp departure from the traditional summative assessment, which evaluates the knowledge of students at the end of the course through a score that summarizes their learning outcomes. Differently, formative assessment is ongoing throughout the teaching of the course, and it evaluates the specific knowledge and skills of the students by assigning multidimensional skill profiles to them. In addition, it informs about the effectiveness of the educational intervention in promoting specific learning. The results of the assessment are then used for planning further steps of teaching and learning.

The theoretical framework of the model is knowledge space theory, a novel and significant approach for the assessment of knowledge introduced by Doignon and Falmagne (1985). Knowledge space theory is fundamentally different from the traditional approach, which is based on the numerical evaluation of some "aptitude". In fact, knowledge space theory provides a non-numerical yet precise representation of the knowledge of students in a given domain. In this way, knowledge space theory is consistent with the aims of formative assessment.

The GaLoM assesses the knowledge of students in the different steps of the learning process, and the effectiveness of the educational intervention in promoting specific learning. The specific features of the model make it particularly suitable for didactic practice.

First of all, the model focuses on the specific skills that the students must possess in order to solve the problems. This works to the advantage of the educational process because it allows teachers to explain theoretically the observed responses, and to predict responses on another collection of problems. Moreover, teachers are helped in identifying which skills should be taught so that the students become able to solve problems that previously they were not able to solve.

In addition, the model is characterized by parameters which provide information relevant at different levels of didactic practice. From one hand, the initial probabilities of the skills, and the gain and loss probabilities provide the teachers with diagnostic information for planning the didactic interventions and evaluating their effectiveness. This information can be obtained at both classroom and student levels. At the classroom level, initial probabilities of the skills enable the teachers to identify what the classroom already knows and what it is ready to learn next. Gain and loss parameters provide program evaluative information. They inform the teachers about the effectiveness of the didactic interventions that have been carried out, and provide them with an objective criterion for choosing, among the didactic tools which are available, the best one for the specific needs and weaknesses of the classroom. This tool should be effective both for learning new skills and for consolidating the skills that the students already possess. Initial probabilities of the skills, and gain and loss probabilities that are obtained for each student, provide the teachers with information relevant to the assessment of the knowledge of each student and the selection of the best educational intervention for his specific weaknesses. Information which derives from these parameters is therefore useful to obtain a detailed investigation of the knowledge and to plan educational interventions which are effective and tailored on the needs of the classroom and individual students. From the other hand, the careless error and lucky guess parameters provide the teachers with information relevant to the validation of their theoretical assumptions and hypotheses about the cognitive processes involved in the response processes, and the validation of the assessment instruments that they use for assessing the knowledge of students. Information which derives from these parameters is therefore useful to realize assessments which are trustworthy and accurate.

The GaLoM was the subject of investigation at different levels. The functioning of the model was analyzed in different conditions of level of information and noise in the data, effects of the learning objects on the skills, and misspecifications of the skill multimap. Moreover, theoretical developments of the model were proposed, which improve its usefulness and informative power in practical applications. The investigations were conducted through simulated studies and empirical applications. Both kinds of studies turned out to be very informative. The former allowed the investigation of the functioning of the model in situations in which all information about the data which is relevant for the analysis is present. In our case, it essentially concerns the skill multimap, the level of noise in the data, and the true value of the parameters. The latter allowed the investigation of the functioning of the model when dealing with the elements of uncertainty that characterize the use of the model in practice. In fact, in practical applications it is not possible to be sure which skills are measured by the assessment instrument and how they are related to the problems, but only assumptions can be made. At the same time, the level of noise in the data is also unknown. The main results of the studies are now reviewed, and some suggestions for increasing the usefulness of the GaLoM which derive from them are presented.

Response data should provide enough information about the skills and they should not be too noisy so that stable and trustworthy estimates of the parameters can be obtained. When this is not the case, compensations among the parameters might result in multiple solutions for their estimates. However, an important role in identifiability is also played by the specific way in which the problems and the skills are related to each other. This is a result that comes out from the entire collection of studies that have been presented, although a systematic investigation of it has not been made yet.

From these results it is possible to draw up strategies that the user can implement in order to increase the chance that model will be identifiable. First, some devices should be used during data collection that reduce the occurrence of noise in the data, such as designing the problems to be open response or the use of assessment without strict time limits. Second, it is important for the user to select collections containing an adequate number of problems in relation to the quantity of skills to be assessed. Not only is the number of problems per skill relevant, it is also fundamental that the combinations of skills associated with each problem are varied and, likewise, situations should be avoided where information concerning a particular skill is derived from only one or two problems.

The model is capable of reproducing the different effects of the learning objects on the skills that might be observed in educational practice. However, there are essentially three elements which might have a detrimental effect on the estimates of the parameters. The first is represented by the level of noise in the data. The accuracy of the estimates decreases as the noise increases. Once again, it is worth stressing the importance of using strategies to reduce such noise. The second is represented by the values of initial probabilities of the skills, and gain and loss probabilities. It is difficult for the model to reproduce a very high gain of a skill when its initial probability is very high, just as it is difficult to reproduce a very high loss of a skill when its initial probability is very low. It should be noted, however, that in such situations the estimation of gain and loss probabilities would not make sense. The third element is represented by the specification of the skill multimap. The incorrect association of the skills to some problems causes an overestimation of the careless error and lucky guess probabilities of those problems. In particular, the omission of a relevant skill from a problem leads to an overestimation of its careless error probabilities, where the inclusion of an irrelevant skills leads to an overestimation of its lucky guess probabilities. These effects have been previously described in Rupp and Templin (2008), and de la Torre and Douglas (2008). Unusually large values of careless error and lucky guess probabilities for some problems might therefore point to the misspecification of the skill multimap for these problems. However, caution is required in interpreting such values as they could merely represent noise in the responses to those particular problems.

The specification of the skill multimap constitutes the core element and is crucial for the assessment to be accurate and trustworthy. The construction of the skill multimap in practical applications should be informed by several sources, including, analysis of the contents of the problems, verbal reports from the students concerning solution strategies utilized and, of course, the knowledge of experts in the domain under investigation. The combined use of these varied sources of information is advisable in constructing the skill multimap. However, it is certainly possible for more than one skill multimap to emerge, each being more or less plausible. This is especially likely when there is not much theory in the domain under investigation. In this case, the best skill multimap has to be selected. The extension of the GaLoM proposed was very useful for this purpose. In particular, the best skill multimaps were identified by comparing the fit to the data of alternative models which underwent a constrained estimation of their careless error and lucky parameters. This approach takes into account the information about model fit which derives from both the standard fit statistics and the estimates of the error parameters. The integration of both types of information facilitates the correct identification of the best skill multimap. The studies presented suggest that the approach is very powerful because it allows the identification of the best skill multimaps from alternatives which differed only in terms of the skills assigned to a few problems. Further work, however, is needed to investigate the viability of the approach in varied conditions. This might involve systematically manipulating characteristics of the skill multimap, misspecifications of the skill multimap (in terms of degree and type), and levels of noise.

In the GaLoM, information relevant for assessing the learning process derives from the gain and loss parameters. These parameters enable the teacher to assess the effectiveness of the educational interventions in promoting specific skills, and to select the best educational intervention from a number of available interventions. A reparametrization of the GaLoM has been proposed that decomposes the gain and loss parameters into a number of logistic parameters. These parameters actually represent a new reading of the information concerning the learning process that can highlight particular features of the skills and of the educational interventions considered. In other words, they provide information in a form that can be used to compare skills without reference to each educational intervention and, conversely, to compare educational interventions without referring to each skill. This might be especially useful when evaluating and comparing a large number of skills and educational interventions. Another important feature of these logistic parameters is that their statistical significance can be tested. The information which derives from these parameters can inform the identification of the features from the various educational interventions which best facilitate the acquisition of new skills, consolidate already possessed skills, and ensure that skills are not lost with time. Subsequently, these same features can be combined to create new and more effective educational interventions available for use by the teacher.

It should be emphasized that the model, and its extensions, can be applied to each knowledge domain in which it is possible to identify a collection of problems and a collection of discrete skills underlying them. The approach proposed in this work offers significant advantages. First, it allows the teacher to reach an accurate assessment of students' knowledge and to develop educational interventions which are effective and tailored to the needs of the students. Moreover, estimates of initial probabilities of the skills, and gain and loss probabilities, which are obtained for each student, can be used to develop individualized learning paths. Second, it allows adaptive procedures to be used for efficient assessment of knowledge; these adaptive procedures have been developed in the context of knowledge space theory (see, e.g., Falmagne et al., 2006; Falmagne & Doignon, 1988a, 1988b). At the core of such procedures is the performance structure. In particular, the specific dependencies between the problems it defines can be used to develop an efficient procedure for the assessment of knowledge, which allows the teacher to reach an accurate assessment of the student by only presenting a subset of the problems.

The present work has been completed with a view to providing teachers with practical tools and procedures that can support teaching practice. Some relevant contributions have been made in this sense. However, further work is necessary in order to increase the usefulness of the proposed model and its applicability to a broad number of contexts. The primary aim of future work is to address the weaknesses of the model in its current form. Some of these will now be discussed.

The model in the current specification assumes the skills to be stochastically independent. This assumption was reasonable for the specific kind of problems that were used in the empirical applications. However, in other contexts it may be unrealistic. The assumption of independence among the skills is not a necessary condition for the model to be applied. In particular, given the probabilities of the competence states at the pretest, any form of dependence among the skills could be conceived. Some such pathways might be log-linear models (see, e.g., Maris, 1999) or Bayesian networks (see, e.g., Mislevy & Gitomer, 1996).

The proposed model can be applied to learning processes with two assessment steps. In real applications, however, it might be useful to allow multiple testing points throughout the teaching of a course. The model in the current form represents a good starting point from which to move forward towards a situation in which any number of assessment steps can be considered. An extension in this direction would qualify the model as a hidden Markov model (see, e.g., Rabiner & Juang, 1986).

Finally, the model does not take into account that students may need certain skills to understand the learning objects they are presented with. This is an important aspect since it might have an impact on the estimation of the gain probabilities. An extension of the model might take into account whether the competence state of the students does or does not contain the prerequisites for understanding a given learning object. One such possibility in this sense consists in associating with each learning object a subset of skills which constitute a prerequisite for dealing with that learning object (see, e.g., Hockemeyer, 2003; Heller, Steiner, Hockemeyer, & Albert, 2006).

The results of this work would certainly increase the usefulness of the proposed approach and its applicability to a broad number of contexts and situations. This would contribute to making it more appealing to teachers who wish to use it as a support for didactic practice.

Appendix A

MATLAB codes

A.1 Code for estimating and testing the GaLoM

The MATLAB function `GaLoM` computes EM estimates of the parameters of the Gain-Loss Model as well as fit indexes.

The function requires as input `pat`, `freq`, `itemskill`, `tol`, `maxiter`, `view`. `pat` is a structure containing m binary matrices with dimensions $l \times 2n$ (where m is the number of learning objects presented to m groups of students, l is the number of different response patterns observed in each group of students, 2 is the number of assessment steps, n is the number of problems). Responses at the pretest precede them at the posttest. `freq` is a structure containing m column vectors of the l observed frequencies of each response pattern. `itemskill` is a $h \times (|S| + 1)$ matrix (where $|S|$ is the number of skills and h is a number that ranges from n to $n \times |2^S \setminus \{\emptyset\}|$) specifying the competences associated with each problem. The first column of the matrix specifies the number of the problem, the others specify if it is associated (1) or not (0) with a skill. `tol` (optional) is the tolerance criterion for the EM algorithm as maximum adjustment among all the model parameters. The default value is 10^{-5}. `maxiter` is a scalar value that specifies the maximum number of iterations in the EM algorithm. The default value is 500. `view` (optional) is a boolean variable indicating if an interim model fit should be displayed (true) or not (false). The default value is "true".

The function returns a structure containing the estimates of the model parameters (pi, gamma, lambda, alpha, beta), and statistics of model fit (chi-square, log-likelihood, AIC, AICC, BIC).

```
function model=GaLoM(pat,freq,itemskill,tol,maxiter,view)

% GaLoM: Compute EM estimates of parameters and fit statistics of
%        Gain-Loss Model

if nargin<6, view=true; end    % Display interim results (true/false)
if nargin<5, maxiter=500; end  % Number of iterations by default
if nargin<4, tol=1e-5; end     % Tolerance value by default
if nargin<3
```

```
    error('Insufficient number of arguments.');
end
if maxiter<1
    error('The number of iterations must be positive.')
end

% Derivate performance states corresponding to skill multimap
[v,w]=skillmap(itemskill);

ngroups=length(pat);    % Number of groups
nstates=size(w,1);      % Number of performance states
nskills=size(w,2);      % Number of skills
nitems=size(v,2);       % Number of problems

grp_sz=zeros(ngroups,1);
gpat=[];
gfreq=[];
for g=1:ngroups
    gpat=[gpat;pat{g}];
    gfreq=[gfreq;freq{g}];
    grp_sz(g)=sum(freq{g});
end
[pat0,freq0]=patfreq(gpat,gfreq);
sz=sum(freq0);

npar=2*nitems+(2*ngroups+1)*nskills;
df=2^(2*nitems)-npar-1;
chi=zeros(maxiter,1);

% Initial values of estimates
alpha=rand(nitems,1)*.5;           % Careless error
beta=rand(nitems,1)*.5;            % Lucky guess
pi=rand(nskills,1);                % Initial probabilities of skills
gamma=rand(nskills,ngroups);       % Gain
lambda=rand(nskills,ngroups);      % Loss

% Main loop of EM algorithm
for iter=1:maxiter

    % Expectation step

    % Compute vector of initial probabilities of states
    p=exp(w*log(pi)+(1-w)*log(1-pi)); % [nstates,1]
    p_rep=repmat(p,1,nstates); % [nstates,nstates]

    % Iterate on groups
    for g=1:ngroups
        pat1=pat{g}(:,1:nitems); % [npat(g),nitems]
        pat2=pat{g}(:,nitems+1:2*nitems); % [npat(g),nitems]

        % Compute matrix of conditional probabilities of response patterns
```

```
    % given states at pretest
    p1=rho(alpha,beta,v,pat1); % [npat(g),nstates]

    % Compute matrix of conditional probabilities of response patterns
    % given states at posttest
    p2=rho(alpha,beta,v,pat2); % [npat(g),nstates]

    % Compute transition probabilities among states
    tau=rho(lambda(:,g),gamma(:,g),w,w); % [nstates,nstates]

    % Compute conditional probabilities of states given response patterns
    b{g}=zeros(nstates,nstates,size(pat1,1));
    for i=1:size(pat1,1)
        b{g}(:,:,i)=(p1(i,:)'*p2(i,:)).*tau.*p_rep; % [nstates,nstates]
        b{g}(:,:,i)=b{g}(:,:,i)/sum(sum(b{g}(:,:,i))); % [nstates,nstates]
    end
end

% Compute Pearson's chi-square for current iteration
ft=sum(expfreq(pat0,grp_sz,alpha,beta,pi,gamma,lambda,v,w),2);
chi(iter)=sum(freq0.*freq0./ft)-sz;
pval=1-gammainc(chi(iter)/2,df/2);

% Display interim fit if requested
if view
fprintf('\nIteration %d of %d. Chi-square %f',iter,maxiter,chi(iter));
figure(1);
semilogy(1:iter,chi(1:iter),'b');
xlabel(sprintf('Iteration %d of %d',iter,maxiter));
ylabel('Chi-square');
title(sprintf('Gain-Loss Model. Chi-square = %g. df = %g. p-value = %g',...
...chi(iter),df,pval));
end

% Maximization step

% Save previous values of estimates
alpha_old=alpha;
beta_old=beta;
gamma_old=gamma;
lambda_old=lambda;
pi_old=pi;

a12=zeros(nitems,1);
b12=a12;
da=a12;
db=b12;
pp=zeros(nskills,1);
% Iterate on groups
for g=1:ngroups
```

```
    % Divide response patterns observed in a group into pretest and
    %posttest
    pat1=pat{g}(:,1:nitems); % [npat(g),nitems]
    pat2=pat{g}(:,nitems+1:2*nitems); % [npat(g),nitems]

    % Compute conditional probabilities of states at pretest
    % given observed response patterns
    q1=sum(permute(b{g},[1,3,2]),3); % [states,npat(g)]

    % Compute conditional probabilities of states at posttest
    % given observed response patterns
    q2=sum(permute(b{g},[2,3,1]),3); % [nstates,npat(g)]

    % Useful in following computations
    qq=(q1+q2)*freq{g}; % [nstates,1]

    % Interim amount for adjustment of alpha
    a1=(q1'*v).*(1-pat1); % [npat(g),nitems]
    a2=(q2'*v).*(1-pat2); % [npat(g),nitems]
    a12=a12+(a1+a2)'*freq{g}; % [nitems,1]
    da=da+v'*qq; % [nitems,1]

    % Interim amount for adjustment of beta
    b1=(q1'*(1-v)).*pat1; % [npat(g),nitems]
    b2=(q2'*(1-v)).*pat2; % [npat(g),nitems]
    b12=b12+(b1+b2)'*freq{g}; % [nitems,1]
    db=db+(1-v)'*qq; % [nitems,1]

    % Compute expected number of pairs (K,L) of states given
    % observed response patterns
    bg=zeros(nstates,nstates);
    for i=1:size(pat1,1);
        bg=bg+freq{g}(i)*b{g}(:,:,i); % [nstates,nstates]
    end
    bg1=sum(bg,2); % [nstates,1]

    % Adjust estimates of gamma and lambda
    gamma(:,g)=sum((1-w).*(bg*w))'./((1-w)'*bg1); % [nskills,1]
    lambda(:,g)=sum(w.*(bg*(1-w)))'./(w'*bg1); % [nskills,1]

    % Interim amount for adjustment of pi
    pp=pp+w'*bg1; % [nskills,1]
end

% Adjust estimates of alpha, beta and pi
alpha=a12./da;
beta=b12./db;
pi=pp/sz;

% Check if tolerance value is reached and, if so, end the loop
alpha_max=max(abs(alpha_old-alpha));
```

```
    beta_max=max(abs(beta_old-beta));
    pi_max=max(abs(pi_old-pi));
    lambda_max=max(max(abs(lambda_old-lambda)));
    gamma_max=max(max(abs(gamma_old-gamma)));
    xmax=max([alpha_max,beta_max,pi_max,lambda_max,gamma_max]);
    if xmax<=tol
        fprintf('\n\nTolerance value reached in GaLoM\n\n');
        break;
    end
end

if iter==maxiter & xmax>tol
    fprintf('\n\nThe estimates do not converge.');
    fprintf('\nIncrease the number of iterations or');
    fprintf('\nthe tolerance value.\n\n');
end

% Statistics of model fit (Pearson's chi-square, log-likelihood,
% AIC, AICC, BIC)
model.ft=sum(expfreq(pat0,grp_sz,alpha,beta,pi,gamma,lambda,v,w),2);
model.chisquare=sum(freq0.*freq0./model.ft)-sz;
model.pvalue=1-gammainc(chi(iter)/2,df/2);
model.loglike=-sum(freq0.*log(model.ft/sz));
model.df=df;
model.aic=2*(model.loglike+npar);
model.aicc=model.aic+2*npar*(npar+1)/(sz-npar-1);
model.bic=2*model.loglike+npar*log(sz);

% Save estimates of model parameters
model.alpha=alpha;
model.beta=beta;
model.pi=pi;
model.gamma=gamma;
model.lambda=lambda;

function [v,w]=skillmap(itemskill)

% skillmap: Derivation of performance states corresponding to skill multimap

nl=max(itemskill(:,1));
ns=size(itemskill,2)-1;
w=powerset(ns);
nk=size(w,1);
v=zeros(nk,nl);
for k=1:nk
    for i=1:size(itemskill,1)
        x=itemskill(i,2:ns+1);
        if (x&w(k,:))==x
            v(k,itemskill(i,1))=1;
        end
    end
```

```
end

function c = rho(a,b,states,patterns)

% rho: Conditional probabilities of response patterns given
% performance states


[m,n] = size(patterns);
eps = 1e-9;
a = a+(a < eps)*eps-(1-a < eps)*eps;   % prevents log of zero
b = b+(b < eps)*eps-(1-b < eps)*eps;
u = ones(m,1);
xa = u*log(a)';
ya = u*log(1-a)';
xb = u*log(b)';
yb = u*log(1-b)';
p = (patterns.*ya)*states';
q = (patterns.*xb)*(1-states)';
r = ((1-patterns).*xa)*states';
s = ((1-patterns).*yb)*(1-states)';
c = exp(p+q+r+s);

function freq=expfreq(pat,sz,alpha,beta,pi,gamma,lambda,v,w)

% expfreq: Compute expected frequencies of Gain-Loss Model

nitems=size(pat,2)/2;
ngroups=size(gamma,2);

pat1=pat(:,1:nitems);
pat2=pat(:,nitems+1:end);

p=exp(w*log(pi)+(1-w)*log(1-pi));
p1=rho(alpha,beta,v,pat1);
p2=rho(alpha,beta,v,pat2);

freq=zeros(size(pat,1),ngroups);
for g=1:ngroups
    tau=rho(lambda(:,g),gamma(:,g),w,w);
    freq(:,g)=sz(g)*(p1.*(p2*tau))*p;
end
```

A.2 Code for computing the probabilities for individual students

The MATLAB function `InProb` computes probabilities of states at pretest and posttest, initial and final probabilities of the skills, gain and loss probabilities for individual students.

The function requires as input `pat`, `model`. `pat` is the same structure containing the observed response patterns used as input of the function `GaLoM`. `model` is the same structure returned by the function `GaLoM`.

The function returns a structure containing the probabilities of states at pretest and posttest (q1, q2), the initial and final probabilities of the skills (ps1, ps2), and the gain and loss probabilities of the skills (gain, loss) for individual students.

```
function individual=InProb(pat,model)

% InProb: Compute probabilities of states at pretest and posttest,
% initial and final probabilities of the skills, gain and loss
% probabilities for individual students



% Model parameters
pi=model.pi;
v=model.v;
w=model.w;
alpha=model.alpha;
beta=model.beta;
gamma=model.gamma;
lambda=model.lambda;

ngroups=length(pat);     % number of groups
nstates=size(w,1);       % number of performance states
nskills=size(w,2);       % number of skills
nitems=size(v,2);        % number of problems

% Compute vector of initial probabilities of states
% (equal for all groups)
p=exp(w*log(pi)+(1-w)*log(1-pi)); % [nstates,1]
p_rep=repmat(p,1,nstates); % [nstates,nstates]
```

```
% Iterate on groups
for g=1:ngroups

    % Divide response patterns observed in a group into
    % pretest and posttest
    pat1=pat{g}(:,1:nitems); % [npat(g),nitems]
    pat2=pat{g}(:,nitems+1:2*nitems); % [npat(g),nitems]


    % Compute matrix of conditional probabilities of response patterns
    % given states at pretest
    p1=rho(alpha,beta,v,pat1); % [npat(g),nstates]

    % Compute matrix of conditional probabilities of response patterns
    % given states at posttest
    p2=rho(alpha,beta,v,pat2); % [npat(g),nstates]

    % Compute transition probabilities among states
    tau=rho(lambda(:,g),gamma(:,g),w,w); % [nstates,nstates]

    % Compute conditional probabilities of states given response patterns
    q12{g}=zeros(nstates,nstates,size(pat1,1));
    for i=1:size(pat1,1)
      q12{g}(:,:,i)=(p1(i,:)'*p2(i,:)).*tau.*p_rep; % [nstates,nstates]
      q12{g}(:,:,i)=q12{g}(:,:,i)/sum(sum(q12{g}(:,:,i))); % [nstates,nstates]
    end


    % Compute conditional probabilities of states at pretest
    % given observed response patterns
    q1{g}=sum(permute(q12{g},[1,3,2]),3); % [states,npat(g)]

    % Compute conditional probabilities of states at pretest
    % given observed response patterns
    q2{g}=sum(permute(q12{g},[2,3,1]),3); % [nstates,npat(g)]

    % Compute conditional probabilities of skills at pretest
    % given observed response patterns
    ps1{g}=(q1{g})'*w; % [npat(g),nskills]

    % Compute conditional probabilities of skills at posttest
    % given observed response patterns
    ps2{g}=(q2{g})'*w; % [npat(g),nskills]

    % Compute gain and loss probabilities of skills
    % given observed response patterns
    for i=1:size(pat1,1)
        for s=1:nskills
            ygain=(1-w(:,s))*w(:,s)';
            yloss=w(:,s)*(1-w(:,s)');
```

```
            gain{g}(i,s)=sum(sum(q12{g}(:,:,i).*ygain))/(1-ps1{g}(i,s));
            loss{g}(i,s)=sum(sum(q12{g}(:,:,i).*yloss))/(ps1{g}(i,s));
        end
    end
end

individual.q1=q1;
individual.q2=q2;
individual.ps1=ps1;
individual.ps2=ps2;
individual.gain=gain;
individual.loss=loss;


function c = rho(a,b,states,patterns)

% rho: Conditional probabilities of response patterns given
% performance states


[m,n] = size(patterns);
eps = 1e-9;
a = a+(a < eps)*eps-(1-a < eps)*eps;   % prevents log of zero
b = b+(b < eps)*eps-(1-b < eps)*eps;
u = ones(m,1);
xa = u*log(a)';
ya = u*log(1-a)';
xb = u*log(b)';
yb = u*log(1-b)';
p = (patterns.*ya)*states';
q = (patterns.*xb)*(1-states)';
r = ((1-patterns).*xa)*states';
s = ((1-patterns).*yb)*(1-states)';
c = exp(p+q+r+s);
```

A.3 Code for estimating and testing the CoGaLoM

The MATLAB function `CoGaLoM` computes EM estimates of the parameters of the Constrained Gain-Loss Model as well as fit indexes.

The function requires as input `pat`, `freq`, `itemskill`, `amax`, `bmax`, `tol`, `maxiter`, `view`. `amax` and `bmax` are two column vectors of length n (where n is the number of problems). They respectively specify the upper bounds of the careless error and lucky guess parameters of each problem. The other arguments of the function correspond to those described in the function `GaLoM`.

The function returns a structure containing the estimates of the model parameters (pi, gamma, lambda, alpha, beta), and statistics of model fit (chi-square, log-likelihood, AIC, AICC, BIC). The estimates of the alpha and beta parameters are constrained.

```
function model=CoGaLoM(pat,freq,itemskill,amax,bmax,tol,maxiter,view)

% CoGaLoM: Compute EM estimates of parameters and fit statistics of
% Constrained Gain-Loss Model

if nargin<8, view=true; end    % Display interim results (true/false)
if nargin<7, maxiter=500; end  % Number of iterations by default
if nargin<6, tol=1e-5; end     % Tolerance value by default
if nargin<5
    error('Insufficient number of arguments.');
end
if maxiter<1
    error('The number of iterations must be positive.')
end

% Derivate knowledge states corresponding to skill multimap
[v,w]=skillmap(itemskill);

ngroups=length(pat);    % Number of groups
nstates=size(w,1);      % Number of knowledge states
nskills=size(w,2);      % Number of skills
nitems=size(v,2);       % Number of problems

amu=ones(size(pat{1},2)/2,1);
```

```
bmu=amu;
rate=0.9;

grp_sz=zeros(ngroups,1);
gpat=[];
gfreq=[];
for g=1:ngroups
    gpat=[gpat;pat{g}];
    gfreq=[gfreq;freq{g}];
    grp_sz(g)=sum(freq{g});
end
[pat0,freq0]=patfreq(gpat,gfreq);
sz=sum(freq0);

npar=2*nitems+(2*ngroups+1)*nskills;
df=2^(2*nitems)-npar-1;
chi=zeros(maxiter,1);

% Initial values of estimates
alpha=rand(nitems,1)*.5;          % Careless error
beta=rand(nitems,1)*.5;           % Lucky guess
pi=rand(nskills,1);               % Initial probabilities of skills
gamma=rand(nskills,ngroups);      % Gain
lambda=rand(nskills,ngroups);     % Loss

% Main loop of EM algorithm
for iter=1:maxiter

    % Expectation step

    % Compute vector of initial probabilities of states
    p=exp(w*log(pi)+(1-w)*log(1-pi)); % [nstates,1]
    p_rep=repmat(p,1,nstates); % [nstates,nstates]

    % Iterate on groups
    for g=1:ngroups
        pat1=pat{g}(:,1:nitems); % [npat(g),nitems]
        pat2=pat{g}(:,nitems+1:2*nitems); % [npat(g),nitems]

        % Compute matrix of conditional probabilities of response patterns
        % given states at pretest
        p1=rho(alpha,beta,v,pat1); % [npat(g),nstates]

        % Compute matrix of conditional probabilities of response patterns
        % given states at posttest
        p2=rho(alpha,beta,v,pat2); % [npat(g),nstates]

        % Compute transition probabilities among states
        tau=rho(lambda(:,g),gamma(:,g),w,w); % [nstates,nstates]

        % Compute conditional probabilities of states given response patterns
```

```
    b{g}=zeros(nstates,nstates,size(pat1,1));
    for i=1:size(pat1,1)
        b{g}(:,:,i)=(p1(i,:)'*p2(i,:)).*tau.*p_rep; % [nstates,nstates]
        b{g}(:,:,i)=b{g}(:,:,i)/sum(sum(b{g}(:,:,i))); % [nstates,nstates]
    end
end

% Compute Pearson's chi-square for current iteration
ft=sum(expfreq(pat0,grp_sz,alpha,beta,pi,gamma,lambda,v,w),2);
chi(iter)=sum(freq0.*freq0./ft)-sz;
pval=1-chi2cdf(chi(iter),df);

% Display interim fit if requested
if view
fprintf('\nIteration %d of %d. Chi-square %f',iter,maxiter,chi(iter));
figure(1);
semilogy(1:iter,chi(1:iter),'b');
xlabel(sprintf('Iteration %d of %d',iter,maxiter));
ylabel('Chi-square');
title(sprintf('Gain-Loss Model. Chi-square = %g. df = %g. p-value = %g',...
...chi(iter),df,pval));
pause(1e-6);
end

% Maximization step

% Save previous values of estimates
alpha_old=alpha;
beta_old=beta;
gamma_old=gamma;
lambda_old=lambda;
pi_old=pi;

a12=zeros(nitems,1);
b12=a12;
da=a12;
db=b12;
pp=zeros(nskills,1);
% Iterate on groups
for g=1:ngroups

    % Divide response patterns observed in a group into pretest and
    % posttest
    pat1=pat{g}(:,1:nitems); % [npat(g),nitems]
    pat2=pat{g}(:,nitems+1:2*nitems); % [npat(g),nitems]

    % Compute conditional probabilities of states at pretest
    % given observed response patterns
    q1=sum(permute(b{g},[1,3,2]),3); % [states,npat(g)]

    % Compute conditional probabilities of states at posttest
```

```
    % given observed response patterns
    q2=sum(permute(b{g},[2,3,1]),3); % [nstates,npat(g)]

    % Useful in following computations
    qq=(q1+q2)*freq{g}; % [nstates,1]

    % Interim amount for adjustment of alpha
    a1=(q1'*v).*(1-pat1); % [npat(g),nitems]
    a2=(q2'*v).*(1-pat2); % [npat(g),nitems]
    a12=a12+(a1+a2)'*freq{g}; % [nitems,1]
    da=da+v'*qq; % [nitems,1]

    % Interim amount for adjustment of beta
    b1=(q1'*(1-v)).*pat1; % [npat(g),nitems]
    b2=(q2'*(1-v)).*pat2; % [npat(g),nitems]
    b12=b12+(b1+b2)'*freq{g}; % [nitems,1]
    db=db+(1-v)'*qq; % [nitems,1]

    % Compute expected number of pairs (K,L) of states given
    % observed response patterns
    bg=zeros(nstates,nstates);
    for i=1:size(pat1,1);
        bg=bg+freq{g}(i)*b{g}(:,:,i); % [nstates,nstates]
    end
    bg1=sum(bg,2); % [nstates,1]

    % Adjust estimates of gamma and lambda
    gamma(:,g)=sum((1-w).*(bg*w))'./((1-w)'*bg1); % [nskills,1]
    lambda(:,g)=sum(w.*(bg*(1-w)))'./(w'*bg1); % [nskills,1]

    % Interim amount for adjustment of pi
    pp=pp+w'*bg1; % [nskills,1]
end

% Adjust estimates of alpha, beta and pi
acoef=2*amu+amax.*(da+amu)+a12;
alpha=(acoef-sqrt(acoef.*acoef-4*amax.*(da+2*amu).*(a12+amu)))./(2*(da+2*amu));
bcoef=2*bmu+bmax.*(db+bmu)+b12;
beta=(bcoef-sqrt(bcoef.*bcoef-4*bmax.*(db+2*bmu).*(b12+bmu)))./(2*(db+2*bmu));
pi=pp/sz;

amu=rate*amu;
bmu=rate*bmu;


% Check if tolerance value is reached and, if so, end the loop
alpha_max=max(abs(alpha_old-alpha));
beta_max=max(abs(beta_old-beta));
pi_max=max(abs(pi_old-pi));
lambda_max=max(max(abs(lambda_old-lambda)));
gamma_max=max(max(abs(gamma_old-gamma)));
```

```
    xmax=max([alpha_max,beta_max,pi_max,lambda_max,gamma_max]);
    if xmax<=tol
        fprintf('\n\nTolerance value reached in CoGaLoM\n\n');
        break;
    end


end

if iter==maxiter & xmax>tol
    fprintf('\n\nThe estimates do not converge.');
    fprintf('\nIncrease the number of iterations or');
    fprintf('\nthe tolerance value.\n\n');
end

% Statistics of model fit (Pearson's chi-square, log-likelihood,
% AIC, AICC, BIC)
model.ft=sum(expfreq(pat0,grp_sz,alpha,beta,pi,gamma,lambda,v,w),2);
model.chisquare=sum(freq0.*freq0./model.ft)-sz;
model.pvalue=1-chi2cdf(chi(iter),df);
model.loglike=-sum(freq0.*log(model.ft/sz));
model.df=df;
model.aic=2*(model.loglike+npar);
model.aicc=model.aic+2*npar*(npar+1)/(sz-npar-1);
model.bic=2*model.loglike+npar*log(sz);

% Save estimates of model parameters
model.alpha=alpha;
model.beta=beta;
model.pi=pi;
model.gamma=gamma;
model.lambda=lambda;

function [v,w]=skillmap(itemskill)

% skillmap: Derivation of knowledge states corresponding to skill multimap

nl=max(itemskill(:,1));
ns=size(itemskill,2)-1;
w=powerset(ns);
nk=size(w,1);
v=zeros(nk,nl);
for k=1:nk
    for i=1:size(itemskill,1)
        x=itemskill(i,2:ns+1);
        if (x&w(k,:))==x
            v(k,itemskill(i,1))=1;
        end
    end
end
```

```
function c = rho(a,b,states,patterns)

% rho: Conditional probabilities of response patterns given
% knowledge states


[m,n] = size(patterns);
eps = 1e-9;
a = a+(a < eps)*eps-(1-a < eps)*eps;   % prevents log of zero
b = b+(b < eps)*eps-(1-b < eps)*eps;
u = ones(m,1);
xa = u*log(a)';
ya = u*log(1-a)';
xb = u*log(b)';
yb = u*log(1-b)';
p = (patterns.*ya)*states';
q = (patterns.*xb)*(1-states)';
r = ((1-patterns).*xa)*states';
s = ((1-patterns).*yb)*(1-states)';
c = exp(p+q+r+s);

function freq=expfreq(pat,sz,alpha,beta,pi,gamma,lambda,v,w)

% expfreq: Compute expected frequencies of Gain-Loss Model

nitems=size(pat,2)/2;
ngroups=size(gamma,2);

pat1=pat(:,1:nitems);
pat2=pat(:,nitems+1:end);

p=exp(w*log(pi)+(1-w)*log(1-pi));
p1=rho(alpha,beta,v,pat1);
p2=rho(alpha,beta,v,pat2);

freq=zeros(size(pat,1),ngroups);
for g=1:ngroups
    tau=rho(lambda(:,g),gamma(:,g),w,w);
    freq(:,g)=sz(g)*(p1.*(p2*tau))*p;
end
```

A.4 Code for computing the logistic parameters

The MATLAB function LogPar decomposes the gain and loss parameters into the logistic parameters.

The function requires as input gamma, lambda. gamma and lambda are two $|S| \times m$ matrices (where S is the number of skills and m is the number of learning objects), which respectively specify the gain and the loss probabilities.

The function returns two structures (lngam, lnlam), which contain the logistic parameters derived from the gain parameters (lngam.phi, lngam.omega, lngam.psi) and from the loss parameters (lnlam.phi, lnlam.omega, lnlam.psi).

```
function [lngam,lnlam]=LogPar(gamma,lambda)

% LogPar: Decompose the estimates of gain and loss parameters into
% logistic parameters

ngroups=size(gamma,2);     % Number of groups

% Decomposition of gain parameters
gr=log(gamma./(1-gamma));
lngam.phi=-mean(gr,2);
lngam.psi=gr+repmat(lngam.phi,1,ngroups);
lngam.omega=mean(lngam.psi);

% Decomposition of loss parameters
lr=log(lambda./(1-lambda));
lnlam.phi=mean(lr,2);
lnlam.psi=-lr+repmat(lnlam.phi,1,ngroups);
lnlam.omega=mean(lnlam.psi);
```

Appendix B

Materials used in the empirical applications

B.1 Materials from Chapter 2

Table B.1: Collection of Problems

Problem	Original	Translated
1	Dato un evento A in uno spazio campionario S, la probabilitá di A é $P(A) = .28$. Trovare $P(\overline{A})$.	Given an event A in a sample space S, the probability of A is $P(A) = .28$. Find $P(\overline{A})$.
2	Dati due eventi A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A \cap B) = .13$; $P(A \cap \overline{B}) = .34$. Trovare $P(A)$.	Given two events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .13$; $P(A \cap \overline{B}) = .34$. Find $P(A)$.
3	Dati due eventi A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A \cap B) = .26$; $P(B) = .50$. Trovare $P(A\|B)$.	Given two events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .26$; $P(B) = .50$. Find $P(A\|B)$.
4	Dati due eventi indipendenti A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A) = .10$; $P(B) = .78$. Trovare $P(A \cap B)$.	Given two independent events A and B in a sample space S, the following probabilities are known: $P(A) = .10$; $P(B) = .78$. Find $P(A \cap B)$.
5	Dati due eventi A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A \cap B) = .86$; $P(A \cap \overline{B}) = .02$. Trovare $P(\overline{A})$.	Given two events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .86$; $P(A \cap \overline{B}) = .02$. Find $P(\overline{A})$.
6	Dati due eventi A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A \cap B) = .08$; $P(\overline{B} = .20)$. Trovare $P(A\|B)$.	Given two events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .08$; $P(\overline{B} = .20)$. Find $P(A\|B)$.
7	Dati due eventi indipendenti A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(\overline{A}) = .95$; $P(B) = .84$. Trovare $P(A \cap B)$.	Given two independent events A and B in a sample space S, the following probabilities are known: $P(\overline{A}) = .95$; $P(B) = .84$. Find $P(A \cap B)$.
8	Dati due eventi A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A \cap B) = .24$; $P(A \cap \overline{B}) = .12$. Trovare $P(B\|A)$.	Given two events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .24$; $P(A \cap \overline{B}) = .12$. Find $P(B\|A)$.
9	Dati due eventi indipendenti A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A \cap B) = .15$; $P(A \cap \overline{B}) = .18$. Trovare $P(B)$.	Given two independent events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .15$; $P(A \cap \overline{B}) = .18$. Find $P(B)$.
10	Dati due eventi indipendenti A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A) = .56$; $P(B) = .78$. Trovare $P(A\|B)$.	Given two independent events A and B in a sample space S, the following probabilities are known: $P(A) = .56$; $P(B) = .78$. Find $P(A\|B)$.
11	Dati due eventi A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(\overline{A}) = .04$; $P(B\|A) = .67$. Trovare $P(A \cap \overline{B})$.	Given two events A and B in a sample space S, the following probabilities are known: $P(\overline{A}) = .04$; $P(B\|A) = .67$. Find $P(A \cap \overline{B})$.
12	Dati due eventi indipendenti A e B in uno spazio campionario S, si conoscono le seguenti probabilitá: $P(A \cap B) = .34$; $P(A \cap \overline{B}) = .18$. Trovare $P(\overline{A} \cap \overline{B})$.	Given two independent events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .34$; $P(A \cap \overline{B}) = .18$. Find $P(\overline{A} \cap \overline{B})$.
13	Dati due eventi indipendenti A e B in uno spazio campionario S, si conosce la seguente probabilitá: $P(A\|B) = .02$. Trovare $P(\overline{A})$.	Given two independent events A and B in a sample space S, the following probability is known: $P(A\|B) = .02$. Find $P(\overline{A})$.

B.2 Materials from Chapter 4

The material is here provided that was used in the empirical application described in Chapter 4. The original and translated versions of the learning object are given in the next section. The collections of problems presented at pretest and posttest are respectively in Tables B.2 and B.3.

Learning object

Original version

Wahrscheinlichkeit für ein Ereignis

Wenn alle Elementarereignisse für die Ergebnismenge S; die gleiche Wahrscheinlichkeit haben, dann ist die Wahrscheinlichkeit, dass ein Ereignis A eintritt:

$$P(A) = \frac{\text{Anzahl der Elemente in } A}{\text{Anzahl aller Elemente in } S}$$

Beispiel: In einer Schachtel befinden sich 10 Murmeln in den folgenden Farben: 6 weiße, 3 rote und 1 schwarze. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Murmel rot ist?

$$P(\text{rot}) = \frac{\text{Anzahl roter Murmeln}}{\text{Anzahl aller Murmeln}} = \frac{3}{10}$$

Komplementärereignis

Wenn A ein Ereignis ist, dann ist die Wahrscheinlichkeit für das Komplementärereignis $P(\overline{A})$:

$$P(\overline{A}) = 1 - P(A)$$

Beispiel: Die Wahrscheinlichkeit aus einem Skatspiel einen Buben zu ziehen sei .125. Wie großist die Wahrscheinlichkeit keinen Buben zu ziehen?

$$P(\text{kein Bube}) = 1 - P(\text{Bube}) = 1 - .125$$

Vereinigung disjunkter Ereignisse

Wenn A und B zwei disjunkte (sich gegenseitig ausschließende) Ereignisse sind, dann ist die Wahrscheinlichkeit ihrer Vereinigung $P(A \cup B)$:

$$P(A \cup B) = P(A) - P(B)$$

Beispiel: Ein Skatspiel besteht aus 32 verschiedenen Karten. Die Wahrscheinlichkeit eine 7 zu ziehen ist .125 und die Wahrscheinlichkeit eine 8 zu ziehen ist ebenfalls .125. Wie großist die Wahrscheinlichkeit eine 7 oder eine 8 zu ziehen?

$$P(7 \text{ oder } 8) = P(7) + P(8) = .125 + .125$$

Unabhängigkeit von Ereignissen

Die Ereignisse A und B sind genau dann unabhngig, wenn die Wahrscheinlichkeit für die Schnittmenge $P(A \cap B)$:

$$P(A \cap B) = P(A)P(B)$$

Beispiel: In einer Bibliothek ist die Wahrscheinlichkeit zufällig ein Buch aus der Kategorie "Roman" auszusuchen .15 und die Wahrscheinlichkeit ein Buch mit rotem Umschlag auszusuchen .30. Angenommen das Ereignis "Roman" ist unabhängig vom Ereignis "roter Umschlag", wie großist dann die Wahrscheinlichkeit für das Ereignis "Roman mit rotem Umschlag"?

$$P(\text{Roman mit rotem Umschlag}) = P(\text{Roman})P(\text{roter Umschlag}) = .15 \times .30$$

Translated version

Determining the probability of an event

If all elementary events in the sample space S have the same probability, the probability of any event A is given by:

$$P(A) = \frac{\text{number of elementary events in } A}{\text{total number of elementary events is } S}$$

Example: A box contains 10 marbles with the following colours: 6 white, 3 red, 1 black. What is the probability that a randomly drawn marble is red?

$$P(\text{red}) = \frac{\text{number of red marbles}}{\text{total number of marbles}} = \frac{3}{10}$$

Complement of events

If A is an event in the sample space S, the probability of its complement $P(\overline{A})$ is given by:

$$P(\overline{A}) = 1 - P(A)$$

Example: The probability of drawing a Jack from a deck is .08. What is the probability of not drawing a Jack?

$$P(\text{not Jack}) = 1 - P(\text{Jack}) = 1 - .08$$

Union of mutually exclusive events

If A and B are two mutually exclusive events in the sample space S, the probability of their union $P(A \cup B)$ is given by:

$$P(A \cup B) = P(A) + P(B)$$

Example: Given a standard deck containing 52 different cards, the probability of drawing a 4 is .08, and the probability of drawing a 5 is .08. What is the probability of drawing a 4 or a 5?

$$P(4 \text{ o } 5) = P(4) + P(5) = .08 + .08$$

Independence of events

If events A and B are independent, the joint probability $P(A \cap B)$ is equal to:

$$P(A \cap B) = P(A)P(B)$$

Example: In a library, the probability of randomly selecting a "fiction" book is .15 and the probability of selecting a book having a red cover is .30. If the event "fiction" is independent of the event "red cover", what is the probability of the event "fiction and red cover"?

$$P(\text{fiction and red cover}) = P(\text{fiction})P(\text{red cover}) = .15 \times .30$$

Table B.2: Collection of Problems presented at the Pretest

Problem	Original	Translated
1	In einer Schachtel befinden sich 30 Murmeln in den folgenden Farben: 8 rote, 10 schwarze und 12 gelbe. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Murmel gelb ist?	A box contains 30 marbles with the following colours: 8 red, 10 black and 12 yellow. What is the probability that a randomly drawn marble is yellow?
2	In einem Beutel befinden sich 5-Cent-, 10-Cent- und 20-Cent-Stücke. Die Wahrscheinlichkeit zufällig ein 5-Cent-Stück zu ziehen ist .35, ein 10-Cent-Stück zu ziehen .25 und die Wahrscheinlichkeit ein 20-Cent-Stück zu ziehen ist .40. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Münze nicht ein 5-Cent-Stück ist?	A bag contains 5-Cent, 10-Cent and 20-Cent pieces. The probability of drawing a 5-Cent piece is .35, that of drawing a 10-Cent piece is .25 and that of drawing a 20-Cent piece is .40. What is the probability that the coin randomly drawn is not a 5-Cent piece?
3	In einem Beutel befinden sich 5-Cent-, 10-Cent- und 20-Cent-Stücke. Die Wahrscheinlichkeit zufällig ein 5-Cent-Stück zu ziehen ist .20, ein 10-Cent-Stück zu ziehen .45 und die Wahrscheinlichkeit ein 20-Cent-Stück zu ziehen ist .35. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Münze ein 5-Cent-Stück oder ein 20-Cent-Stück ist?	A bag contains 5-Cent, 10-Cent and 20-Cent pieces. The probability of drawing a 5-Cent piece is .20, that of drawing a 10-Cent piece is .45 and that of drawing a 20-Cent piece is .35. What is the probability that the coin randomly drawn is a 5-Cent or a 20-Cent piece?
4	An einer Schule sind 40% der Schüler Jungen und 80% der Schüler Rechtshänder. Angenommen Geschlecht und Händigkeit sind voneinander unabhängig. Wie großist die Wahrscheinlichkeit zufällig einen rechtshändigen Jungen auszusuchen?	In a school the probability of observing a male pupil is 40% and the probability of observing a right-handed pupil is 80%. Assume that the hand writing is independent from the gender. What is the probability of observing a pupil that is male and right-handed?
5	Sie haben ein Skatspiel mit 32 Karten. Wie großist die Wahrscheinlichkeit kein Herz zu ziehen?	Consider a standard deck containing 32 different cards. What is the probability of not drawing a heart?
6	In einer Schachtel befinden sich 20 Murmeln in den folgenden Farben: 4 weiße, 14 grüne und 2 rote. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Murmel nicht weißist?	A box contains 20 marbles with the following colours: 4 white, 14 green and 2 red. What is the probability that a randomly drawn marble in not white?
7	In einer Schachtel befinden sich 10 Murmeln in den folgenden Farben: 2 gelbe, 5 blaue und 3 rote. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Murmel gelb oder blau ist?	A box contains 10 marbles with the following colours: 2 yellow, 5 blue and 3 red. What is the probability that a randomly drawn marble is yellow or blue?
8	Wie großist die Wahrscheinlichkeit mit einem Würfel eine gerade Zahl zu werfen?	What is the probability of obtain an even number by throwing a dice?
9	Sie haben ein Skatspiel mit 32 Karten. Wie großist die Wahrscheinlichkeit eine 9 in einer schwarzen Farbe zu ziehen?	Consider a standard deck containing 32 different cards. What is the probability of drawing a 9 in a black suit?
10	In einer Schachtel befinden sich rote und gelbe Murmeln, die großoder klein sein können. Die Wahrscheinlichkeit eine rote Murmel zu ziehen ist .70, die Wahrscheinlichkeit eine kleine Murmel zu ziehen ist .40. Angenommen die Farbe der Murmeln ist unabhängig von ihrer Größe. Wie großist die Wahrscheinlichkeit zufällig eine kleine Murmel zu ziehen, die nicht rot ist?	A box contains marbles that are red or yellow, small or large. The probability of drawing a red marble is .70, the probability of drawing a small marble is .40. Assume that the colour of the marbles is independent from their size. What is the probability of randomly drawing a marble that is yellow and small?
11	In einer Garage befinden sich 50 Autos. Davon sind 20 Autos schwarz und 10 Autos tanken Diesel. Angenommen Farbe und Art der Tankfüllung sind voneinander unabhängig. Wie großist die Wahrscheinlichkeit, dass ein zufällig ausgesuchtes Auto nicht schwarz ist und Diesel tankt?	In a garage there are 50 cars. 20 are black and 10 are diesel. Assume that the colour of the cars is independent from the fuel. What is the probability that a car is not black and it is diesel?
12	In einer Schachtel befinden sich 20 Murmeln. 10 Murmeln sind rot, 6 sind gelb und 4 sind schwarz. 12 Murmeln sind klein und 8 Murmeln sind groß. Angenommen die Farbe der Murmeln ist unabhängig von ihrer Größe. Wie großist die Wahrscheinlichkeit zufällig eine kleine Murmel zu ziehen, die gelb oder rot ist?	A box contains 20 marbles. 10 marbles are red, 6 are yellow and 4 are black. 12 marbles are small and 8 are large. Assume that the colour of the marbles is independent from their size. What is the probability of randomly drawing a marble that is red or yellow and that is small?

Note. The 32 card deck corresponds to a typical 52 card deck with no values from 2 to 6. It is popular in Germany.

Table B.3: Collection of Problems presented at the Posttest

Problem	Original	Translated
1	In einer Schachtel befinden sich 30 Murmeln in den folgenden Farben: 10 rote, 14 gelbe und 6 grüne. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Murmel grün ist?	A box contains 30 marbles with the following colours: 10 red, 14 yellow and 6 green. What is the probability that a randomly drawn marble is green?
2	In einem Beutel befinden sich 5-Cent-, 10-Cent- und 20-Cent-Stücke. Die Wahrscheinlichkeit zufällig ein 5-Cent-Stück zu ziehen ist .25, ein 10-Cent-Stück zu ziehen .60 und die Wahrscheinlichkeit ein 20-Cent-Stück zu ziehen ist .15. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Münze nicht ein 5-Cent-Stück ist?	A bag contains 5-Cent, 10-Cent and 20-Cent pieces. The probability of drawing a 5-Cent piece is .25, that of drawing a 10-cent piece is .60 and that of drawing a 20-Cent piece is .15. What is the probability that the coin randomly drawn is not 5-Cent piece?
3	In einem Beutel befinden sich 5-Cent-, 10-Cent- und 20-Cent-Stücke. Die Wahrscheinlichkeit zufällig ein 5-Cent-Stück zu ziehen ist .35, ein 10-Cent-Stück zu ziehen .20 und die Wahrscheinlichkeit ein 20-Cent-Stück zu ziehen ist .45. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Münze ein 5-Cent-Stück oder ein 20-Cent-Stück ist?	A bag contains 5-Cent, 10-Cent and 20-Cent pieces. The probability of drawing a 5-Cent piece is .35, that of drawing a 10-Cent piece is .20 and that of drawing a 20-Cent piece is .45. What is the probability that the coin randomly drawn is a 5-Cent or or a 20-Cent piece?
4	An einer Schule sind 70% der Schl̈er Mädchen und 10% der Schüler Linkshänder. Angenommen Geschlecht und Händigkeit sind voneinander unabhängig. Wie großist die Wahrscheinlichkeit zufällig ein linkshändiges Mädchen auszusuchen?	In a school the probability of observing a female pupil is 70% and the probability of observing a left-handed pupil is 10%. Assume that the hand writing is independent from the gender. What is the probability of observing a pupil that is female and left-handed?
5	Sie haben ein Skatspiel mit 32 Karten. Wie großist die Wahrscheinlichkeit kein Kreuz zu ziehen?	Consider a standard deck containing 32 different cards. What is the probability of not drawing a club?
6	In einer Schachtel befinden sich 20 Murmeln in den folgenden Farben: 6 gelbe, 10 rote und 4 grüne. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Murmel nicht gelb ist?	A box contains 20 marbles with the following colours: 6 yellow, 10 red, 4 green. What is the probability that a randomly drawn marble in not yellow?
7	In einer Schachtel befinden sich 10 Murmeln in den folgenden Farben: 5 blaue, 3 rote und 2 grüne. Wie großist die Wahrscheinlichkeit, dass eine zufällig gezogene Murmel blau oder rot ist?	A box contains 10 marbles with the following colours: 5 blue, 3 red, 2 green. What is the probability that a randomly drawn marble is blue or red?
8	Wie großist die Wahrscheinlichkeit mit einem Würfel eine ungerade Zahl zu werfen?	What is the probability of obtaining an odd number by throwing a dice?
9	Sie haben ein Skatspiel mit 32 Karten. Wie großist die Wahrscheinlichkeit eine 10 in einer roten Farbe zu ziehen?	Consider a standard deck containing 52 different cards. What is the probability of drawing a 10 in a red suit?
10	In einer Schachtel befinden sich grüne und rote Murmeln, die großoder klein sein knnen. Die Wahrscheinlichkeit eine grüne Murmel zu ziehen ist .40, die Wahrscheinlichkeit eine große Murmel zu ziehen ist .20. Angenommen die Farbe der Murmeln ist unabhängig von ihrer Größe. Wie großist die Wahrscheinlichkeit zufällig eine große Murmel zu ziehen, die nicht grün ist?	A box contains marbles that are green or red, large or small. The probability of drawing a green marble is .40, the probability of drawing a large marble is .20. Assume that the colour of the marbles is independent from their size. What is the probability of randomly drawing a marble that is red and large?
11	In einer Garage befinden sich 50 Autos. Davon sind 15 Autos weißund 20 Autos tanken Diesel. Angenommen Farbe und Art der Tankfüllung sind voneinander unabhängig. Wie großist die Wahrscheinlichkeit, dass ein zufällig ausgesuchtes Auto nicht weißist und Diesel tankt?	In a garage there are 50 cars. 15 are white and 20 are diesel. Assume that the colour of the cars is independent from the fuel. What is the probability that a car is not white and it is diesel?
12	In einer Schachtel befinden sich 20 Murmeln. 8 Murmeln sind weiß, 4 sind grün und 8 sind rot. 15 Murmeln sind klein und 5 Murmeln sind groß. Angenommen die Farbe der Murmeln ist unabhängig von ihrer Größe. Wie großist die Wahrscheinlichkeit zufällig eine große Murmel zu ziehen, die weißoder grün ist?	A box contains 20 marbles. 8 marbles are white, 4 are green and 8 are red. 5 marbles are large and 15 are small. Assume that the colour of the marbles is independent from their size. What is the probability of randomly drawing a marble that is white or green and that is large?

Note. The 32 card deck corresponds to a typical 52 card deck with no values from 2 to 6. It is popular in Germany.

B.3 Materials from Chapter 5

The material is here provided that was used in the empirical application described in Chapter 5. The good learnin objetc is given in the next section in the original and translated version. The bad learning object correspond to the good learning object without the examples. The two collections of problems denoted by form A and form B are respectively in Tables B.4 and B.5.

Good Learning object

Original version

Determinare la probabilità di un evento

Se tutti gli esiti nello spazio campionario S sono equiprobabili, la probabilità di un qualunque evento A è data da:

$$P(A) = \frac{\text{numero di esiti in } A}{\text{numero totale di esiti in } S}$$

Esempio: Una scatola contiene 10 palline dei seguenti colori: 6 bianche, 3 rosse, 1 nera. Qual è la probabilità che una pallina estratta a caso sia rossa?

$$P(\text{rossa}) = \frac{\text{numero di palline rosse}}{\text{numero totale di palline}} = \frac{3}{10}$$

Complemento di eventi

Se A è un evento nello spazio campionario S, la probabilità del suo complemento $P(\overline{A})$ è data da:

$$P(\overline{A}) = 1 - P(A)$$

Esempio: Dato un mazzo contenente 52 diverse carte da gioco, la probabilità di estrarre un Jack è $\frac{4}{52}$. Qual è la probabilità di estrarre una carta che non sia un Jack?

$$P(\text{no Jack}) = 1 - P(\text{Jack}) = 1 - \frac{4}{52}$$

Unione di eventi incompatibili

Se A e B sono due eventi incompatibili nello spazio campionario S, la probabilità della loro unione $P(A \cup B)$ è data da:

$$P(A \cup B) = P(A) + P(B)$$

Esempio: Dato un mazzo contenente 52 diverse carte da gioco, la probabilità di estrarre un 4 è $\frac{4}{52}$, e la probabilità di estrarre un 5 è $\frac{4}{52}$. Qual è la probabilità di estrarre un 4 o un 5?

$$P(4 \text{ o } 5) = P(4) + P(5) = \frac{4}{52} + \frac{4}{52}$$

Probabilità congiunta di eventi indipendenti

Se A e B sono due eventi indipendenti, la loro probabilità congiunta $P(A \cap B)$ è data da:

$$P(A \cap B) = P(A) \times P(B)$$

Esempio: Si lanci un dado due volte. Qual è la probabilità di ottenere un 2 al primo lancio e un 4 al secondo lancio?

$$P(2 \text{ primo lancio} \cap 4 \text{ secondo lancio}) =$$
$$P(2 \text{ primo lancio}) \times P(4 \text{ secondo lancio}) = \frac{1}{6} \times \frac{1}{6}$$

Alcuni problemi richiedono solo uno dei quattro concetti presentati per essere risolti, altri richiedono pi di un concetto. Rispetto a questi ultimi, è importante individuare i concetti necessari ed utilizzarli nell'ordine appropriato.

Esempio: Dato un mazzo contenente 52 diverse carte da gioco, qual è la probabilità di estrarre un 5 o un 7?

Passo 1: Si calcola la probabilità di estrarre un 5 e quella di estrarre un 7.

$$P(5) = \frac{\text{numero di carte con il 5}}{\text{numero totale di carte}} = \frac{4}{52}$$
$$P(7) = \frac{\text{numero di carte con il 7}}{\text{numero totale di carte}} = \frac{4}{52}$$

Passo 2: Si calcola la probabilità di estrarre un 5 o un 7.

$$P(5 \text{ o } 7) = P(5) + P(7) = \frac{4}{52} + \frac{4}{52}$$

Translated version

Determining the probability of an event

If all the outcomes in sample space S are equiprobable, the probability of given event A is given by:

$$P(A) = \frac{\text{number of outcomes in } A}{\text{total number of outcomes in } S}$$

Example: A box contains 10 marbles of the following colours: 6 white, 3 red, 1 black. What is a probability of a marble, extracted randomly, being red?

$$P(\text{red}) = \frac{\text{number of red marbles}}{\text{total number of marbles}} = \frac{3}{10}$$

Complement of events

If A is an event in sample space S, the probability of its complement $P(\overline{A})$ is given by:

$$P(\overline{A}) = 1 - P(A)$$

Example: Assuming a deck of cards contains 52 different playing cards, the probability of extracting a Jack is $\frac{4}{52}$. What is the probability of extracting a card which is not a Jack?

$$P(\text{no Jack}) = 1 - P(\text{Jack}) = 1 - \frac{4}{52}$$

Union of mutially exclusive events

If A and B are two mutually exclusive events in sample space S, the probability of their union $P(A \cup B)$ is given by:

$$P(A \cup B) = P(A) + P(B)$$

Example: Assuming a deck of cards contains 52 different playing cards, the probability of extracting a 4 is $\frac{4}{52}$, and the probability of extracting a 5 is $\frac{4}{52}$. What is the probability of extracting a 4 or a 5?

$$P(4 \text{ or } 5) = P(4) + P(5) = \frac{4}{52} + \frac{4}{52}$$

Joint probability of independent events

If A and B are two independent events, their joint probability $P(A \cap B)$ is given by:

$$P(A \cap B) = P(A) \times P(B)$$

Esempio: One die is thrown twice. What is the probability of obtaining a 2 on the first throw and a 4 on the second throw?

$$P(2 \text{ first throw} \cap 4 \text{ second throw}) = P(2 \text{ first throw}) \times P(4 \text{ second throw}) = \frac{1}{6} \times \frac{1}{6}$$

Some problems only require one of the 4 concepts presented in order to be resolved, other problems require more than one concept. With regard to the these problems, it is important to identify the necessary concepts and to use them in the correct order.

Example: Assuming a deck of cards contains 52 different playing cards, what is the probability of extracting a 5 or a 7?

Step 1: Calculation of the probability of extracting a 5 and the probability of extracting a 7.

$$P(5) = \frac{\text{number of cards with 5}}{\text{total number of cards}} = \frac{4}{52}$$
$$P(7) = \frac{\text{number of cards with 7}}{\text{total number of cards}} = \frac{4}{52}$$

Step 2: Calculation of the probability of extracting a 5 or a 7.

$$P(5 \text{ or } 7) = P(5) + P(7) = \frac{4}{52} + \frac{4}{52}$$

Table B.4: Collection of Problems of Form A

Problem	Original	Translated
1	Una scatola contiene palline verdi, bianche e rosse. La probabilità di estrarre una pallina verde è $\frac{2}{15}$, quella di estrarre una pallina bianca è $\frac{7}{15}$, e quella di estrarre una pallina rossa è $\frac{6}{15}$. Qual è la probabilità che una pallina estratta a caso sia verde o rossa?	A box contains green, white and red marbles. The probability of extracting a green marble is $\frac{2}{15}$, that of extracting a white marble is $\frac{7}{15}$, and that of extracting a red marble is $\frac{6}{15}$. What is the probability of a marble, extracted randomly, being green or red?
2	Dato un mazzo contenente 52 diverse carte da gioco, qual è la probabilità di non estrarre una carta di cuori?	Assuming a deck of cards contains 52 different playing cards, what is the probability of not extracting a hearts card?
3	In un garage ci sono 50 automobili. Di queste, 20 sono nere e 10 sono diesel. Se il colore delle automobili è indipendente dal carburante, qual è la probabilità che un'automobile non sia nera e sia diesel?	In a garage there are 50 cars. Of these, 20 are black and 10 are diesel. If the colour of the car is independent of the fuel type, what is the probability of a car not being black but being diesel?
4	Dato un mazzo contenente 52 diverse carte da gioco, qual è la probabilità di estrarre un 4 di colore nero?	Assuming a deck of cards contains 52 different playing cards, what is the probability of extracting a black 4?
5	Una scatola contiene palline che sono rosse o gialle, piccole o grandi. La probabilità di estrarre una pallina rossa è $\frac{7}{10}$, la probabilità di estrarre una pallina piccola è $\frac{2}{5}$. Se il colore delle palline è indipendente dalla loro grandezza, qual è la probabilità di estrarre a caso una pallina che sia gialla e piccola?	A box contains marbles that are red or yellow, small or large. The probability of extracting a red marble is $\frac{7}{10}$, the probability of extracting a small marble is $\frac{2}{5}$. If the colour of the marbles is independent of their size, what is the probability of randomly extracting a marble which is yellow and small?
6	In una scuola, la probabilità di osservare uno alunno maschio è $\frac{2}{5}$ e la probabilità di osservare un alunno destrimane è $\frac{4}{5}$. Se la mano con cui un alunno scrive è indipendente dal suo genere, qual è la probabilità di osservare un alunno che sia maschio e destrimane?	In a school, the probability of observing a male student is $\frac{2}{5}$ and the probability of observing a right-handed student is $\frac{4}{5}$. If the handedness of a student is independent of gender, what is the probability of observing a student who is male and right-handed?
7	Si lanci un dado. Qual è la probabilità di non ottenere né un 2 né un 5?	Throw a dice. What is the probability of obtaining neither a 2 nor a 5?
8	Una scatola contiene palline che sono rosse, gialle, bianche e nere, piccole e grandi. La probabilità di estrarre una pallina rossa è $\frac{7}{25}$, quella di estrarre una pallina gialla è $\frac{4}{25}$, quella di estrarre una pallina bianca è $\frac{12}{25}$, e quella di estrarre una pallina nera è $\frac{2}{25}$. La probabilità di estrarre una pallina piccola è $\frac{1}{3}$. Se il colore delle palline è indipendente dalla loro grandezza, qual è la probabilità di estrarre a caso una pallina che sia rossa o gialla e che sia piccola?	A box contains marbles which are red, yellow, white, black, small and large. The probability of extracting a red marble is $\frac{7}{25}$, that of extracting a yellow marble is $\frac{4}{25}$, that of extracting a white marble is $\frac{12}{25}$, and that of extracting a black marble is $\frac{2}{25}$. The probability of extracting a small marble is $\frac{1}{3}$. If the colour of the marbles is independent of their size, what is the probability of randomly extracting a marble which is both small and yellow or red?
9	Si lanci un dado cento volte. Qual è la probabilità di ottenere 2 o 3 al primo lancio e di ottenere 4, 5 o 6 al secondo lancio?	Throw a dice 100 times. What is the probability of obtaining 2 or 3 on the first throw and of obtaining 4, 5 or 6 on the second throw?
10	Una scatola contiene 6 palline di cui 4 bianche e 2 nere. Un'altra scatola contiene 8 palline di cui 3 bianche e 5 nere. Si estragga a caso una pallina da ciascuna scatola. Qual è la probabilità che entrambe le palline siano bianche?	A box contains 6 marbles of which 4 are white and 2 are black. Another box contains 8 marbles of which 3 are white and 5 are black. 1 marble is extracted randomly from each box. What is the probability of both marbles being white?
11	Si lanci un dado. Qual è la probabilità di ottenere un 1 o un 4?	Throw a dice. What is the probability of obtaining a 1 or a 4?
12	In una biblioteca la probabilità di prendere un libro di narrativa è $\frac{3}{20}$ e la probabilità di prendere un libro con la copertina rossa è $\frac{3}{5}$. Se il genere letterario del libro è indipendente dal colore della copertina, qual è la probabilità di prendere un libro di narrativa che non abbia la copertina rossa?	In a library the probability of picking up a fiction book is $\frac{3}{20}$ and the probability of picking up a book with a red cover is $\frac{3}{5}$. If the book genre is independent of the colour of the cover, what is the probability of picking up a fiction book which does not have a red cover?
13	Una scatola contiene 75 palline dei seguenti colori: 10 rosse, 20 azzurre, 30 bianche e 15 gialle. Qual è la probabilità che una pallina estratta a caso non sia né rossa né azzurra?	A box contains 75 marbles of the following colours: 10 red, 20 blue, 30 white and 15 yellow. What is the probability of randomly extracting a marble that is neither red nor blue?
14	Una scatola contiene palline rosse, gialle e verdi. La probabilità di estrarre una pallina rossa è $\frac{3}{10}$. Qual è la probabilità di estrarre una pallina non rossa?	A box contains red, yellow and green marbles. The probability of extracting a red marble is $\frac{3}{10}$. What is the probability of extracting a marble that is not red?
15	Si lanci un dado cento volte. Qual è la probabilità di non ottenere 5 al primo lancio e di non ottenere né 1 né 2 al secondo lancio?	Throw a dice 100 times. What is the probability of not obtaining 5 on the first throw and of obtaining neither 1 nor 2 on the second throw?
16	Una scatola contiene 10 palline di cui 3 verdi. Un'altra scatola contiene 20 palline di cui 4 verdi. Si estragga a caso una pallina da ciascuna scatola. Qual è la probabilità che né la pallina estratta dalla prima scatola, né quella estratta dalla seconda siano verdi?	A box contains 10 marbles of which 3 are green. Another box contains 20 marbles of which 4 are green. 1 marble is extracted from each box. What is the probability that neither the first marble, nor the second marble extracted will be green?
17	Una scatola contiene 20 palline dei seguenti colori: 4 bianche, 14 verdi, 2 rosse. Qual è la probabilità che una pallina estratta a caso non sia bianca?	A box contains 20 marbles of the following colours: 4 white, 14 green, 2 red. What is the probability that a randomly extracted marble is not white?
18	Una scatola contiene 20 palline. 10 palline sono rosse, 6 sono gialle e 4 sono nere. 12 palline sono piccole e 8 sono grandi. Se il colore delle palline è indipendente dalla loro grandezza, qual è la probabilità di estrarre a caso una pallina che sia rossa o gialla e che sia piccola?	A box contains 20 marbles. 10 marbles are red, 6 are yellow and 4 are black. 12 marbles are small and 8 are large. If the colour of the marbles is independent of their size, what is the probability of randomly extracting a marble which is small and red or yellow?
19	Una scatola contiene 10 palline dei seguenti colori: 2 gialle, 5 blu, 3 rosse. Qual è la probabilità che una pallina estratta a caso sia gialla o blu?	A box contains 10 marbles of the following colours: 2 yellow, 5 blue, 3 red. What is the probability of a marble, extracted randomly, being yellow or blue?

Table B.5: Collection of Problems of Form B

Problem	Original	Translated
1	Una scatola contiene palline blu, nere e gialle. La probabilità di estrarre una pallina blu $\frac{6}{15}$, quella di estrarre una pallina nera $\frac{2}{15}$, e quella di estrarre una pallina gialla $\frac{7}{15}$. Qual è la probabilità che una pallina estratta a caso sia blu o gialla?	A box contains blue, black and yellow marbles. The probability of extracting a blue marble is $\frac{6}{15}$, that of extracting a black marble is $\frac{2}{15}$, and that of extracting a yellow marble is $\frac{7}{15}$. What is the probability of a marble, extracted randomly, being blue or yellow?
2	Dato un mazzo contenente 52 diverse carte da gioco, qual è la probabilità di non estrarre una carta di fiori?	Assuming a deck of cards contains 52 different playing cards, what is the probability of not extracting a clubs card?
3	In un garage ci sono 50 automobili. Di queste, 15 sono bianche e 20 sono dicesel. Se il colore delle automobili è indipendente dal carburante, qual è la probabilità che unautomobile non sia bianca e sia dicesel?	In a garage there are 50 cars. Of these, 15 are white and 20 are dicesel. If the colour of the car is independent of the fuel type, what is the probability of a car not being white but being dicesel?
4	Dato un mazzo contenente 52 diverse carte da gioco, qual è la probabilità di estrarre un 10 di colore rosso?	Assuming a deck of cards contains 52 different playing cards, what is the probability of extracting a red 10?
5	Una scatola contiene palline che sono blu o verdi, piccole o grandi. La probabilità di estrarre una pallina blu è $\frac{3}{10}$, la probabilità di estrarre una pallina piccola è $\frac{4}{5}$. Se il colore delle palline è indipendente dalla loro grandezza, qual è la probabilità di estrarre a caso una pallina che sia verde e piccola?	A box contains marbles that are blue or green, small or large. The probability of extracting a blue marble is $\frac{3}{10}$, the probability of extracting a small marble is $\frac{4}{5}$. If the colour of the marbles is independent of their size, what is the probability of randomly extracting a marble which is green and small?
6	In una scuola, la probabilità di osservare uno alunno maschio è $\frac{3}{7}$ e la probabilità di osservare un alunno destrimane è $\frac{5}{7}$. Se la mano con cui un alunno scrive è indipendente dal suo genere, qual è la probabilità di osservare un alunno che sia maschio e destrimane?	In a school, the probability of observing a male student is $\frac{3}{7}$ and the probability of observing a right-handed student is $\frac{5}{7}$. If the handedness of a student is independent of gender, what is the probability of observing a student who is male and right-handed?
7	Si lanci un dado. Qual è la probabilità di non ottenere né un 3 né un 4?	Throw a dice. What is the probability of obtaining neither a 3 nor a 4?
8	Una scatola contiene palline che sono verdi, rosse, bianche e nere, piccole e grandi. La probabilità di estrarre una pallina verde è $\frac{8}{25}$, quella di estrarre una pallina rossa è $\frac{6}{25}$, quella di estrarre una pallina bianca è $\frac{9}{25}$, e quella di estrarre una pallina nera è $\frac{2}{25}$. La probabilità di estrarre una pallina piccola è $\frac{2}{3}$. Se il colore delle palline è indipendente dalla loro grandezza, qual è la probabilità di estrarre a caso una pallina che sia verde o rossa e che sia piccola?	A box contains marbles which are green, red, white, black, small and large. The probability of extracting a green marble is $\frac{8}{25}$, that of extracting a red marble is $\frac{6}{25}$, that of extracting a white marble is $\frac{9}{25}$, and that of extracting a black marble is $\frac{2}{25}$. The probability of extracting a small marble is $\frac{2}{3}$. If the colour of the marbles is independent of their size, what is the probability of randomly extracting a marble which is both small and green or red?
9	Si lanci un dado cento volte. Qual è la probabilità di ottenere 1 o 4 al primo lancio e di ottenere 2, 3 o 5 al secondo lancio?	Throw a dice 100 times. What is the probability of obtaining 1 or 4 on the first throw and of obtaining 2, 3 or 5 on the second throw?
10	Una scatola contiene 8 palline di cui 5 gialle e 3 rosse. Unaltra scatola contiene 6 palline di cui 4 gialle e 2 rosse. Si estragga a caso una pallina da ciascuna scatola. Qual è la probabilità che entrambe le palline siano gialle?	A box contains 8 marbles of which 5 are yellow and 3 are red. Another box contains 6 marbles of which 4 are yellow and 2 are red. 1 marble is extracted randomly from each box. What is the probability of both marbles being yellow?
11	Si lanci un dado. Qual è la probabilità di ottenere un 2 o un 6?	Throw a dice. What is the probability of obtaining a 2 or a 6?
12	In una biblioteca la probabilità di prendere un libro di narrativa è $\frac{7}{20}$ e la probabilità di prendere un libro con la copertina blu è $\frac{2}{5}$. Se il genere letterario del libro è indipendente dal colore della copertina, qual è la probabilità di prendere un libro di narrativa che non abbia la copertina blu?	In a library the probability of picking up a fiction book is $\frac{7}{20}$ and the probability of picking up a book with a blue cover is $\frac{2}{5}$. If the book genre is independent of the colour of the cover, what is the probability of picking up a fiction book which does not have a blue cover?
13	Una scatola contiene 75 palline dei seguenti colori: 30 bianche, 10 verdi, 20 rosse e 15 blu. Qual è la probabilità che una pallina estratta a caso non sia né bianca né verde?	A box contains 75 marbles of the following colours: 30 white, 10 green, 20 red and 15 blue. What is the probability of randomly extracting a marble that is neither whie nor green?
14	Una scatola contiene palline bianche, rosse e gialle. La probabilità di estrarre una pallina bianca è $\frac{6}{10}$. Qual è la probabilità di estrarre una pallina non bianca?	A box contains white, red and yellow marbles. The probability of extracting a white marble is $\frac{6}{10}$. What is the probability of extracting a marble that is not white?
15	Si lanci un dado cento volte. Qual è la probabilità di non ottenere 2 al primo lancio e di non ottenere né 4 né 5 al secondo lancio	Throw a dice 100 times. What is the probability of not obtaining 2 on the first throw and of obtaining neither 4 nor 5 on the second throw?
16	Una scatola contiene 10 palline di cui 4 blu. Unaltra scatola contiene 20 palline di cui 5 blu. Si estragga a caso una pallina da ciascuna scatola. Qual è la probabilità che né la pallina estratta dalla prima scatola, né quella estratta dalla seconda siano blu?	A box contains 10 marbles of which 4 are blue. Another box contains 20 marbles of which 5 are blue. 1 marble is extracted from each box. What is the probability that neither the first marble, nor the second marble extracted will be blue?
17	Una scatola contiene 20 palline dei seguenti colori: 6 blu, 2 rosse, 12 gialle. Qual è la probabilità che una pallina estratta a caso non sia blu?	A box contains 20 marbles of the following colours: 6 blue, 2 red, 12 yellow. What is the probability that a randomly extracted marble is not blue?
18	Una scatola contiene 20 palline. 8 palline sono blu, 10 sono bianche e 2 sono gialle. 6 palline sono piccole e 14 sono grandi. Se il colore delle palline è indipendente dalla loro grandezza, qual è la probabilità di estrarre a caso una pallina che sia blu o bianca e che sia piccola?	A box contains 20 marbles. 8 marbles are blue, 10 are white and 2 are yellow. 6 marbles are small and 14 are large. If the colour of the marbles is independent of their size, what is the probability of randomly extracting a marble which is small and blue or white?
19	Una scatola contiene 10 palline dei seguenti colori: 5 blu, 3 rosse, 2 verdi. Qual è la probabilità che una pallina estratta a caso sia blu o rossa?	A box contains 10 marbles of the following colours: 5 blue, 3 red, 2 green. What is the probability of a marble, extracted randomly, being blue or red?

References

Agresti, A. (2002). *Categorical data analysis, 2nd edition.* New York, NY: Wiley.

Agresti, A. (2007). *An introduction to categorical data analysis, 2nd edition.* New York, NY: Wiley.

Agresti, A., & Lang, J. B. (1993). Quasy-symmetric latent class models, with application to rater agreement. *Biometrics*, *49*, 131–139.

Albert, D., & Held, T. (1994). Establishing knowledge spaces by systematical problem construction. In A. D. (Ed.), *Knowledge structures* (pp. 78–111). New York, NY: Springer-Verlag.

Albert, D., & Lukas, J. (Eds.). (1999). *Knowledge spaces: Theories, empirical research, applications.* Mahwah, NJ: Lawrence Erlbaum.

Anselmi, P., Stefanutti, L., & Robusto, E. (Submitted). Parameter identifiability and goodness of recovery in the Constrained Gain-Loss Model.

Arminger, G., Clogg, C. C., & Sobel, M. E. (Eds.). (1995). *Handbook of statistical modeling for the social and behavioral sciences.* New York, NY: Plenum Press.

Clogg, C. C. (1979). Some latent structure models for the analysis of Likert-type data. *Social Sciences Research*, *8*, 287–301.

Cosyn, E., & Thiéry, N. (2000). A practical procedure to build a knowledge structure. *Journal of Mathematical Psychology*, *44*, 383–407.

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*, 343–362.

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353.

de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*, 595–624.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, *39(1)*, 1–38.

DiBello, L. V., & Stout, W. (2007). Guest editors introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, *44*, 285–291.

Doignon, J.-P. (1994). Knowledge spaces and skill assignments. In G. H. Fisher & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 111–121). New York, NY: Springer-Verlag.

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

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces.* Berlin: Springer-Verlag.

Dowling, C. E. (1993). On the irredundant generation of knowledge spaces. *Journal of Mathematical Psychology*, *37*(1), 49–62.

Düntsch, I., & Gediga, G. (1995). Skills and knowledge structures. *British Journal of Mathematical and Statistical Psychology*, *48*, 9–27.

Falmagne, J.-C. (1989). A latent trait theory via a stochastic learning theory for a knowledge space. *Psychometrika*, *54*(2), 283–303.

Falmagne, J.-C. (1993). Stochastic learning paths in a knowledge structure. *Journal of Mathematical Psychology*, *37*, 489–512.

Falmagne, J.-C. (1996). Errata to SLP. *Journal of Mathematical Psychology*, *40*, 169–174.

Falmagne, J.-C., Cosyn, E., Doignon, J.-P., & Thiéry, N. (2006). The assessment of knowledge, in theory and in practice. In B. Ganter & L. Kwuida (Eds.), *Formal Concept Analysis, 4th International Conference, ICFCA 2006, Dresden, Germany, february 13-17, 2006, Lecture Notes in Artificial Intelligence* (pp. 61–79). Berlin: 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.

Falmagne, J.-C., & Doignon, J.-P. (1988b). A markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology*, *32*, 232–258.

Falmagne, J.-C., & Doignon, J.-P. (2011). *Learning spaces: Interdisciplinary applied mathematics.* Berlin: Springer-Verlag.

Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review*, *97*(2), 201–224.

Falmagne, J.-C., & Lakshminarayan, K. (1994). Stochastic learning paths – Estimation and simulation. In G. H. Fisher & D. Laming (Eds.), *Contributions* *to mathematical psychology, psychometrics and methodology* (pp. 91–110). New York, NY: Springer-Verlag.

Fiacco, A. V., & McCormick, G. P. (Eds.). (1990). *Nonlinear programming: Sequential unconstrained minimization techniques.* Philadelphia, PA: SIAM.

Gediga, G., & Düntsch, I. (2002). Skill set analysis in knowledge structures. *British Journal of Mathematical and Statistical Psychology*, *55*, 361–384.

Haberman, S. (1979). *Qualitative data analysis (Vols. 1, 2).* New York, NY: Academic Press.

Held, T., & Korossy, K. (1998). Data analysis as a heuristic for establishing theoretically founded item structures. *Zeitschrift für Psychologie*, *206*(2), 169–188.

Heller, J., Levene, M., Keenoy, K., Albert, D., & Hockemeyer, C. (2007). Cognitive aspects of trails: A stochastic model linking navigation behaviour to the learner's cognitive state. In J. Schoonenboom, J. Heller, K. Keenoy, M. Levene, & M. Turcsanyi-Szabo (Eds.), *Trails in education: Technologies that support navigational learning* (pp. 119–146). Rotterdam: Sense Publisher.

Heller, J., & Repitsch, C. (2008). Distributed skill functions and the meshing of knowledge structures. *Journal of Mathematical Psychology*, *52*, 147–157.

Heller, J., Steiner, C., Hockemeyer, C., & Albert, D. (2006). Competence-based knowledge structures for personalised learning. *International Journal on E-Learning*, *5*, 75–88.

Hockemeyer, C. (2003). Competence based adaptive e-learning in dynamic domains. In F. W. Hesse & Y. Tamura (Eds.), *The joint workshop of cognition and learning through mediacommunication for advanced e-learning (jwcl)* (pp. 79–82). Berlin.

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.

Kambouri, M., Koppen, M., Villano, M., & Falmagne, J.-C. (1994). Knowledge assessment: Tapping human expertise by the QUERY routine. *International Journal of Human-Computer Studies*, *40*, 119–151.

Koppen, M. (1993). Extracting human expertise for constructing knowledge spaces: An algorithm. *Journal of Mathematical Psychology*, *37*, 1–20.

Koppen, M., & Doignon, J.-P. (1990). How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology*, *34*, 311–331.

Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, and applications* (pp. 103–132). Mahwah: Lawrence Erlbaum Associates.

Lakshminarayan, K., & Gilson, F. (1998). An application of a stochastic knowledge structure model. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology, scientific psychology series* (pp. 155–172). Mahwah: Lawrence Erlbaum Associates.

Langeheine, R., Pannekoek, J., & van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research*, *24*, 492–516.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mill.

Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*, 96–107.

Lukas, J., & Albert, D. (1993). Knowledge assessment based on skill assignment and psychological task analysis. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 139–159). Amsterdam: Elsevier Science Publishers B.V.

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational and Behavioral Statistics*, *2*, 99–120.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187–212.

Masters, G. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13–103). New York, NY: Macmillian.

Mislevy, R. J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction*, *5*, 253–282.

Nichols, P., & Sugrue, B. (1999). The lack of fidelity between cognitively complex constructs and conventional test development practice. *Educational Measurement: Issues and Practice*, *18*, 18–29.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Rabiner, L. R., & Juang, B. H. (1986). An introduction to hidden Markov models. *IEEE Acoustics, Speech & Signal Processing, Magazine*, *3*, 4–16.

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.

Rupp, A. A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, *68*, 78–96.

Sargin, A., & Ünlü, A. (2009). Inductive item tree analysis: Corrections, improvements, and comparisons. *Mathematical Social Sciences*, *58*, 376–392.

Schrepp, M. (1999a). Extracting knowledge structures from observed data. *British Journal of Mathematical and Statistical Psychology*, *52*, 213–224.

Schrepp, M. (1999b). On the empirical construction of implications between bi-valued test items. *Mathematical Social Sciences*, *38*, 361–375.

Schrepp, M. (2002). Explorative analysis of empirical data by boolean analysis of questionnaires. *Zeitschrift für Psychologie*, *210*(2), 99–109.

Schrepp, M. (2003). A method for the analysis of hierarchical dependencies between items of a questionnaire. *Methods of Psychological Research*, *19*, 43–79.

Stefanutti, L. (2006). A logistic approach to knowledge structures. *Journal of Mathematical Psychology*, *50*, 545–561.

Stefanutti, L., Anselmi, P., & Robusto, E. (in press). Assessing learning processes with the Gain-Loss Model. *Behavior Research Methods*.

Stefanutti, L., & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74*, 83–96.

Taagepera, M., Potter, F., Miller, G. E., & Lakshminarayan, K. (1997). Mapping students' thinking patterns by the use of the knowledge space theory. *International Journal of Science Education*, *19*, 283–302.

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

von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of Psychological Research*, *2*, 29–48.

Wright, S. J. (Ed.). (1997). *Primal-dual interior-point methods.* Philadelphia, PA: SIAM.