Università degli Studi di Padova
Dipartimento di Psicologia Generale

Scuola di Dottorato in Scienze Psicologiche XXIII Ciclo
Indirizzo Psicologia Sperimentale

**Elaborato Finale**

# Formal Psychological Assessment
# Theoretical and Mathematical Foundations

Direttore della Scuola: *Ch.mo Prof. Clara Casco*

Coordinatore d'indirizzo: *Ch.mo Prof. Umberto Castiello*

Supervisore: *Ch.mo Prof. Giulio Vidotto*

Dottorando: *Andrea Spoto*

*To my father*

# Contents

# Riassunto

Il presente lavoro di tesi raccoglie i risultati teorici e matematici che costituiscono il fondamento per lo sviluppo del *Formal Psychological Assessment* (FPA). Il FPA nasce dall'applicazione di due teorie sviluppatesi nel settore della Psicologia Matematica alla psicologia clinica e, più in particolare, all'assessment psicodiagnostico: la *Knowledge Space Theory* (KST) e la *Formal Concept Analysis* (FCA). I primi due capitoli della tesi sono finalizzati all'introduzione di tutti i concetti di base che poi saranno ripresi e utilizzati negli ultimi tre capitoli quando saranno presentati i risultati formali e clinici del lavoro svolto nel corso dei tre anni.

Nel primo capitolo è descritto lo stato dell'arte riguardante l'assessment psicologico in generale e quello psicodiagnostico in particolare. Particolare attenzione è rivolta alla cosiddetta integrazione verticale (ossia la gerarchia di approfodimenti diagnostici successivi), al colloquio psicologico, alle interviste semi-strutturate (SCID I e II) e alle scale CBA 2.0. Tale focus è giustificato dalle caratteristiche peculiari cui il FPA s'ispira. Infatti, l'obiettivo finale del FPA è di poter realizzare un assessment psicologico di tipo adattivo, capace cioè di operare in maniera simile a un'intervista semi-strutturata e un colloquio, con il vantaggio di applicare un'integrazione verticale frutto d'inferenze logicamente corrette perché derivanti dai modelli matematico-

formali utilizzati per la costruzione dell'assessment stesso. Questo modello di assessment vede il suo anticipatore naturale nella batteria diagnostica CBA 2.0, strumento di riferimento per la diagnostica psicologica cognitivo-comportamentale italiana, i cui autori per primi intuirono la necessità di costruire uno strumento capace non solo di applicare un'etichetta diagnostica a un paziente, ma di fornire una quantità di elementi necessaria alla formulazione del caso stesso in forma il più possibile standardizzata.

Il secondo capitolo del presente lavoro scende nei dettagli formali e matematici della KST e della FCA. Le due teorie sono presentate introducendo sia i concetti teorici di base, sia tutti quegli elementi in seguito impiegati nella formalizzazione del FPA. In particolare, si fa riferimento per la KST ai concetti di *spazio di conoscenza*, *struttura di conoscenza*, *skill map* e *Basic Local Independence Model* (BLIM), mentre per la FCA s'introducono i concetti di *oggetto formale*, *attributo formale*, *contesto formale* e *concetto formale*. Sono infine esplorate le sovrapposizioni fra le due teorie viste nell'ottica di una loro applicazione congiunta alla psicologia clinica. A tal riguardo, è descritto come sia possibile, attraverso alcuni passaggi formali, partire da un contesto formale (derivato da una matrice booleana di oggetti per attributi) ed ottenere uno spazio di conoscenza. Questi ultimi passaggi formali costituiscono la base teorica applicata nel lavoro presentato nel terzo capitolo.

Il terzo capitolo è il primo della parte propositiva del lavoro di tesi. In esso è presentato il risultato dell'applicazione della FCA e della KST a un insieme di item clinici al fine di costruire uno spazio di conoscenza i cui parametri probabilistici sono stati testati facendo riferimento al modello probabilistico BLIM. Il lavoro rappresenta una prima possibile modalità di derivazione del modello di base per il FPA. L'intuizione fondamentale che guida il lavoro introdotto, e che rappresenta il riferimento concettuale fondamentale del FPA, risiede nella possibilità di poter descrivere ciascun disturbo psicologico (inteso come oggetto formale della FCA) come un insieme di sintomi e caratteristiche diagnostiche facenti riferimento a un qualche background teorico (gli attributi formali della FCA). Dunque appare possibile e plausibile costruire un contesto formale avente come oggetti gli item di un questionario che indaga un certo disturbo e, come attributi, le caratteristiche diagnostiche che ciascun item esplora. In questo modo sarà possibile approfondire la conoscenza delle caratteristiche di un rispondente andando oltre il semplice valore numerico ottenuto al test e descrivendo l'insieme di caratteristiche diagnostiche che egli presenta. La possibilità di passare dalla FCA alla KST è cruciale per l'applicazione del BLIM e la validazione del modello ottenuto. Inoltre, proprio dalla validazione del modello si ricavano due indici per ciascun item studiato (probabilità di falso positivo e falso negativo) utili al fine di una riformulazione e una nuova calibrazione del modello.

Nel quarto capitolo della tesi è presentata una seconda possibile modalità di derivazione del modello di base per il FPA. Tale metodologia fa totalmente riferimento alla KST e, in particolare, al concetto di *skill multi map* ottenibile attraverso il *competency model*. In questo caso la matrice di riferimento iniziale, contenente sempre in riga gli item e in colonna gli attributi diagnostici, rappresenta una skill multi map. Nel corso del capitolo sono presentati i vantaggi metodologici di questo tipo di modello. In particolare si approfondiranno: la possibilità di abbassare fino a 0 i parametri di falso positivo e falso negativo per ciascun item, la possibilità di avere più di una possibile configurazione sintomatologica alla base della risposta ad un dato item, la possibilità di individuare attributi specifici per singoli item, la possibilità di individuare e assegnare valori di probabilità ad un numero $n$ di classi latenti, la possibilità di stimare la probabilità della presenza di ciascun attributo in ogni possibile classe latente.

Nell'ultimo capitolo, infine, si presenta un approccio derivante dagli ultimi sviluppi della KST avvenuti presso la ALEKS Corporation. Questo centro di ricerca, che rappresenta sul piano applicativo il punto più alto della KST, si sta muovendo per la costruzione di strutture di conoscenza, dal metodo basato sull'interrogazione di esperti (*expert query*) a un nuovo metodo basato sull'interrogazione di database (*database query*) in cui sono raccolti i dati degli assessment effettuati nel corso degli anni. Tale interrogazione si fonda su di un algoritmo che procede selezionando e testando le relazioni di prerequisito che soddisfino specifici parametri probabilistici. L'algoritmo in questione è stato applicato anche a un insieme di dati raccolti attraverso un questionario clinico (lo stesso utilizzato nei precedenti capitoli). La struttura derivante da tale interrogazione (questa volta completamente scissa da qualsiasi legame con eventuali caratteristiche diagnostiche sottostanti) rappresenta un ulteriore possibile metodo di concretizzazione del FPA.

In conclusione sono discussi gli aspetti di vantaggio e svantaggio per ciascuno degli approcci presentati negli ultimi tre capitoli e si propongono alcuni possibili sviluppi futuri per il FPA.

# Abstract

The present thesis collects both theoretical and mathematical results representing the basis for the development of *Formal Psychological Assessment* (FPA). FPA arises from the application in the clinical psychology context, and, more specifically, in the psychodiagnostic assessment, of two mathematical psychology theories: *Knowledge Space Theory* (KST) and *Formal Concept Analysis* (FCA). The first two chapters of the thesis aim at introducing all the basic concepts further used in the last three chapters when presenting both formal and clinical results of the three years work.

In the first chapter we introduce the state of art in psychological assessment in general and, specifically, in psychodiagnostic assessment. We focus on the so called vertical integration (i.e. the hierarchy of successive diagnostic depths), on clinical interview, on semi-structured interviews (SCID I and II), and on the CBA 2.0 battery. These elements are crucial with respect to FPA. In fact, the main aim of FPA is the construction of an adaptive assessment tool able to work as a semi-structured interview and to provide the clinician with a certainly correct inference procedure based on logical implication. This assessment model has its natural forerunner in the diagnostic battery CBA 2.0 that is a reference point for cognitive behavioral assessment in Italy. The authors of CBA 2.0 were the first psychologists feeling the need of creating a

tool not only able to provide a diagnostic label, but also to collect the crucial elements for a standardized case formulation.

In the second chapter are introduced both formal and mathematical details of KST and FCA. The basic concepts of the two theories will be presented together with those specific elements used in formalizing FPA. More specifically we will refer to concepts like *knowledge space*, *knowledge structure*, *skill map* and *Basic Local Independence Model* (BLIM) for KST; and to concepts like *formal object*, *formal attribute*, *formal context* and *formal concept* for FPA. Finally, we will explore potential overlaps between the two theories in the perspective of their conjoint application in the clinical framework. In this respect, we will show how it is possible, through some formal passages, to derive a knowledge space from a formal context (consisting in a boolean matrix objects $\times$ attributes). These passages represent the theoretical base for the application presented in chapter three.

In the third chapter are presented the results of the application of KST and FCA to a set of clinical items. Such application is aimed at building a knowledge space whose probabilistic parameters have been tested through the BLIM. This chapter presents the first possible methodology to derive the starting point for FPA. The core idea here is the opportunity to describe each psychological disorder (as a formal object in FCA) using a set of symptoms a diagnostic characteristics referred to a specific theoretical background (the formal attributes of FCA). Thus it seems plausible to build a formal context having the items of a questionnaire as objects and the diagnostic criteria investigated by each single item as attributes. In this way it will be possible to better investigate the characteristics of a subject. This information will be much more complete than the mere numeric score of the test. The possibility to shift from FCA to KST is crucial for the application of BLIM and the model validation. Furthermore, out of the validation, it is possible to obtain a set of indexes for each item (i.e. false positive and false negative probabilities) useful in model eventual reformulation and calibration.

In the fourth chapter of this dissertation we introduce a different way to derive the starting point of FPA. This methodology totally refers to KST and, more specifically, to the concept of *skill multi map* obtained through the so called *competency model*. In this case the basic matrix, containing once again items in rows and diagnostic attributes in columns, represents a skill multi map. In the chapter the advantages of this second approach are presented: the possibility to deflate to 0 false positive and false negative parameters for each single item; the possibility to have more than one symptomatic configuration behind the positive answer to a specific item; the possibility to detect specific attributes for single items; the possibility to estimate probability values to $n$ potential latent classes; the possibility to estimate the probability of each single attribute in each single latent class.

Finally, in the last chapter, we introduce a different approach derived from the latest developments carried out at the ALEKS Corporation. In fact, these corporation, representing the worldwide most relevant practical application of KST, is shifting from the *expert query* to the *database query* in building knowledge structures. The employed databases are those collected through the years at the corporation. The interrogation is based on an algorithm that selects and tests the prerequisite relations among items satisfying a set of probabilistic parameters. This algorithm has been applied to a set of clinical items administered to a large sample of individuals (the same used in the previous chapters). The so built structure (this time totally free from any link with eventually underlying diagnostic criteria) represents a further potential method to realize FPA.

In conclusion, advantages and limitations are discussed for all the presented approaches, and some further developments of FPA are suggested.

# Chapter 1

# Psychological Assessment: State of Art

## 1.1 Introduction

In general, assessment can be described as the collection of a certain amount of information referred to a specific object. Assessment can be applied to the most part of human activities, starting from the assessment of knowledge (Falmagne, Doignon, Cosyn, & Thiéry, 2001), up to the more recent crucial role in distance health (Murphy, Levant, Hall, & Gluekauf, 2007). Assessment can represent a preliminary step, an intermediate feedback or a final evaluation. For instance, assessment can be used to verify whether a person could be adequate for a specific role within an organization (preliminary step), an evaluation could be carried out in order to check whether a specific intervention is working correctly or some modifications are needed (intermediate feedback), finally an exam could inform the teacher about the knowledge presented by a student at the end of a course (final evaluation).

Here we will focus on a specific kind of assessment that is psychological assessment. Psychological assessment is crucial to the definition, training and practice of professional psychology (Groth-Marnat, 2009). The most part of clinicians spends up to four hours to carry out the several parts of clinical assessment (Camara, Nathan, & Puente, 2000). There are several definitions of psychological assessment each one centered on a specific characteristic of this wide concept. More specifically, according to Sanavio and Cornoldi (2001) psychological assessment can be defined as the continuous and active process carried out by a clinician in order to evaluate an individual (Wolpe & Turkat, 1985). Different areas can be explored by psychological assessment and several aims can be achieved through this process. For instance, personality traits, clinical disorders and attitudes could be the object of the assessment and these areas could be investigated for several different reasons such as diagnosis formulation, job selection, forensic test.

In the present dissertation we will focus specifically on the psychological assessment which is carried out at the beginning of the therapeutic work and is aimed at collecting useful information for the case formulation. This kind of psychological assessment is usually named clinical assessment.

## 1.2 Clinical Assessment

Clinical assessment is the first action a clinician has to carry out when working with a patient. This action is aimed at collecting diagnostic elements to be used in programming his further intervention (Spoto, Vidotto, Postal, & Pendoni, 2008). The quality of the initial assessment represents one of the main predictors of a good therapeutic outcome. This explains why through the years an increasing amount of research has focused on the development of the more in more accurate tools for clinical assessment (for a complete review see Meyer et al., 2006). In general, clinical assessment, according to Sanavio (2002), is composed by three main levels of information: the subjective level, the behavioral level, and the physiological level. The subjective level includes all the information collectable through the verbal channel, i.e. clinical interviews, questionnaires, diaries and tests. The behavioral level allows the collection of information through the direct observation performed by the clinician on the patient. For instance many clinicians assert that, whenever interested in clearly understanding patient's claustrophobic reactions, a psychologist could take the patient to an elevator and directly observe the behavioral indicators of anxiety and fear. This kind of measure, they affirm, can be much more informative than any questionnaire, test or interview. Another example could be represented by the observation of the eating behavior of a patient with eating disorder. In these cases it is suggested to have a lunch with the patient in order to easily identify his/her real eating behavior. Finally the physiological level includes all those objective observations that can be carried out through physiological registration (e.g. skin conductance, cardiac frequency, temperature). These measures have grown attention and specificity in the last years (Barlow & Durand, 2004) and are widely used in the investigation of specific disorders such as headache and hypertension (Nicassio, Meyerowitz, & Kerns, 2004) and emotional disorders (Barlow, 2002). Thus, the whole clinical assessment includes several different operations beyond merely the administration of test and interviews (Finn, 2007; Garb, 2007).

All of the three levels of information are important and present some critical issues: for instance, a clinical interview is very useful, but it is time expensive, very seldom standardized and it has been shown how clinicians who rely exclusively on interviews are prone to incomplete understandings (Meyer et al., 2006); for a behavioral observation to be reliable the observers have to be adequately trained and the object under observation has to be clearly identified; finally the physiological observations are often considered less efficient and adequate than an unstructured idiographic approach (Groth-Marnat, 2009). Thus, a multilevel approach in clinical assessment is needed (Sanavio, 2002) in order to account for the different areas involved in this procedure. The integration of the three sources of information is generally called *horizontal integration* (Sanavio & Cornoldi, 2001; Sanavio, 2002). A second crucial kind of integration that has to be performed by the clinician through logical inference (i.e. problem solving and decision making), is the so called *vertical integration*. In other words, the clinician is asked to formulate hypotheses and then try to check the correspondence between such hypotheses and the patient. The critical point in this operation is the degree to which the clinician is able to perform a logically correct inference, and how he is able to include all the critical information within such integration (Sanavio, Bertolotti, Michielin, Vidotto, & Zotti, 1986, 1997, 2008).

We are going now to introduce the main investigation tools that can be used by a clinician to perform an assessment. Our attention will be mainly focused on questionnaires and semi-structured interviews.

### 1.2.1 The Clinical Interview

In this section we are going to refer to the first interview, i.e. the first time clinician and patient meet each other. Usually this first interview is not structured that much, it does not exist a specific sequence of questions to be posed, but the sequence is determined by hypothesis and data (Wolpe & Turkat, 1985; Sanavio, 1991). Nevertheless, a topographic theory of the first interview has been proposed (Sanavio, 2007). This approach subdivides the first interview into 9 main parts:

1. Preliminary phases;
2. Problem exposition;
3. Problem Specification;
4. Functional analysis;
5. Related problems;
6. Problem history;
7. Personal history;
8. Expectations and therapeutic proposals;
9. Closure.

It has to be stressed that the amount and reliability of collected information depends both on clinician's ability and patient's insight. The advantages of this method to collect information are several and in someway the interview is one of the indispensable parts of the assessment. Nevertheless, it has to be noticed that the quality of the information could be affected by the memory of the patient, by his degree of trust in the clinician, by the amount of details of the problem he's able to remind, and by the amount of bias the clinician unconsciously introduces in conducting the interview (Galeazzi & Franceschina, 2001). Furthermore, the patient could report false events and memories. Finally, the most critical aspect of the interview is that it is, especially when the psychologist has not a clear idea of how to perform it, time consuming and it relies on the assumption that the inferences carried out by the clinician are correct. Especially this second assumption could be critical conducting to wrong diagnosis and case formulation.

### 1.2.2 The Observation

The observation is the registration and collection of patient's explicit behaviors. The crucial aspect of this investigation tool is twofold: on the one hand the behavior(s) under investigation has to be precisely defined; on the other hand, the observer has to be accurately trained. If one of these two conditions did not occur the observation would be neither accurate nor reliable.

The first of these two issues is usually addressed through standard observational grids present in literature. These grids specify exactly the behavior to be observed and insert such behavior within a theoretical framework. For instance, one of the items included in the Autism Diagnostic Observation Schedule (ADOS; C. Lord et al., 1989), asks the clinician to observe the number of times the eyes of the child are oriented to the psychologist. The way to observe this behavior is clearly described in the scoring rules of the instrument. In the manual it is also explained the reason why it is important to observe such specific behavior in the diagnosis of a autistic spectrum disorder.

In order to cope with the second issue observers has to undergo a specific training. For instance, in order to administer the ADOS a training is required and a special licence has to be achieved.

A particular kind of observation is the self-monitoring. This procedure is crucial for both assessment and therapy (Sanavio, 1991). The fact that the self-monitoring of a behavior could modify the behavior itself is well known (Kazdin, 1974). This characteristic is critical in the assessment phase, but becomes important in during the treatment, in fact, the simple monitoring of a problematic behavior could influence the frequency of its emission (this principle is used, for example, in the smoking treatment).

### 1.2.3 Projective Techniques

Projective tools refer to the assumption that the subject's response are due to the freudian projection mechanism (Frank, 1939). These methods ask the individual to interpret ambiguous stimuli and the psychologist interprets each answer as a projection of patient's way to look at the reality. These methods rose at the beginning of the $20^{th}$ century and they are still used by psychodynamic psychologists. The most popular projective tool is the Rorschach test (Rorschach, 1927). Other instruments are the Thematic Apperception Test (TAT; Murray, 1943), the Children Apperception Test (CAT; Bellak & Bellak, 1949), the Draw-a-Person Test (DAP; Goodenough, 1926), and the Patte Noir Test (PN; Corman, 1969). Limits of these techniques are mainly related to the fact that there are several problems in the interpretation and scoring of answers (different clinicians could assign different scores to the same test; low inter rater reliability) and more in general in the fact that they are not standardized techniques (Ercolani & Perugini, 1997).

### 1.2.4 Structured and Semi-structured Interviews

Structured interviews can be described as verbally administered questionnaires. Structured interviews are standardized and allow the clinician to assign a specific score to set of constructs. One of the most popular tool of this category is the Camberwell Family Interview (Rutter & Brown, 1966). Much more interesting from our standpoint are the semi-structured interviews. In fact, this diagnostic tool is standardized as well as similar to a traditional interview. In other words semi-structured interviews present a set of items, but not all of them have to be asked by the clinician who is guided in selecting the following question given the previous answer. This critical element introduces the concept of adaptivity of the assessment that will be crucial in the Formal Psychological Assessment (FPA; Spoto, Vidotto, & Stefanutti, 2010). The international reference point for semi-structured interviews is the Structured Clinical Interview for DSM IV (SCID). Two main versions of the SCID are available: the first one refers to the investigation of the disorders included in the first axis of DSM IV (SCID-I; First, Spitzer, Gibbon, & Williams, 1996), the second version investigates the disorders of the second axis of DSM IV (SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997). These interviews investigate the main disorders included in the Diagnostic and Statistical Manual of Mental Disorders IV-Text Revision (DSM-IV-TR; American Psychiatric Association [APA], 1995) such as: depression, anxiety disorder, addictive disorders. The main structure of SCID is modular. Every module is independent from the others and investigates a specific diagnostic area. The interview is aimed not only at evaluating the presence/absence of the symptoms, but also at understanding how the patient reports them. The crucial idea of this procedure, that will be used also in FPA, is the fact that the selection of the areas to be investigated depends on the answers collected up to the specific moment. In other words the clinician is guided through a path of questions. Some critical aspects consist in the time consumption of this procedure, and on the fact that, even if guided, the inference process could be affected by logical biases introduced by the clinician. In the last years have grown interest some computer-administered interviews presenting potential positive aspects together with some critical issues mostly related to the artificiality of the so built situation (Garb, 2007).

### 1.2.5 Psychological Testing

Psychological testing is probably one of the most developed topics in clinical assessment. In literature are present a huge number of questionnaires, surveys, check lists, etc. On the basis of the phase of the assessment in which the clinician is, he can use different questionnaires. At the beginning of the assessment wide spectrum tools are needed, in the second phase of the assessment, when several diagnosis have been excluded, the clinician could be interested in going into the details of some specific disorders by using specifically focused tools. Wide spectrum tools are in general time consuming and provide the clinician with a big amount of information about the patient. The most popular wide spectrum tool is the Minnesota Multiphasic Personality Inventory 2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), while in the Italian context the most popular wide spectrum assessment battery is the Cognitive Behavioural Assessment 2.0 (CBA 2.0; Sanavio et al., 1997). Examples of questionnaires investigating specific problems are the following: Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996), the Scale for Interpersonal Behaviour (SIB; Arrindell & VanDerEnde, 1985) and the Padua Inventory (PI; Sanavio, 1988). These three questionnaires investigate three specific areas of clinical interest: BDI investigates the comprehensive construct of depression, SIB investigates the anxiety related to specific behaviors and the probability that an individual has to do those behaviors (assertiveness), PI investigates the characteristics of obsessions and compulsions presented by a subject. It is evident that the administration of one or more of these tools depends on what the clinician supposes about the potential diagnosis of the patient. If the psychologist has already excluded a potential depression in the patient, it becomes useless to administer BDI to that patient.

The critical issues related to the use of clinical tests are several and well explored in literature. In the present work we will focus on the fact that the main result we can obtain out of a questionnaire is a score (a number) that could have been obtained through several response patterns. One of the main characteristic of FPA is the opportunity provided to the clinician to have information about the response pattern actually observed (i.e. its probability, and the probability of false positive and false negative of each single item of that pattern).

In the next section we will focus on CBA 2.0 battery. This assessment tool represents an ideal reference point of FPA and it presents several aspects that will be the starting point of the formalization of FPA.

### 1.2.6 Cognitive Behavioural Assessment 2.0

#### Introduction

The battery was built at the beginning of the eighties in order to address the need of more adequate assessment tools. The main ideas that guided the psychologists who created the CBA were i) to create a non-projective assessment tool, and ii) to optimize the investigation by reducing the administration, scoring and storage times. Given these issues the battery resulted extremely flexible and usable both in a clinical context, and in the normal population for research aims. Furthermore, the structure of the battery allows for the collection of general information useful to orient further investigations through the use of more specific questionnaires. In fact, CBA 2.0

has both an horizontal and a vertical extension. The horizontal one includes the 10 sheets of the "primary scales", the vertical one depends on the scores of the primary scales and includes a set of specific tools investigating well determined disorders. The tools included in this last set are called "secondary scales" (Sanavio et al., 1986, 1997, 2008; Sanavio, 2002).

### Primary Scales

Primary scales are aimed at providing the clinician with a wide spectrum picture of the patient. They investigate the main psychological disorders and include an anamnestic sheet useful to collect information on a really wide set of areas. The primary scales of the CBA 2.0 are ten, namely:

1. *Sheet* 1: general and personal data; it includes 25 items useful to collect useful practical information.

2. *Sheet* 2: State Trait Anxiety Inventory X1 (STAI-X1); it is the state anxiety scale of the Italian version of Spielberger's questionnaire (Spielberger, Gorsuch, & Lushene, 1970; Lazzari & Pancheri, 1980). This questionnaire is aimed at testing the state anxiety at the beginning of the test. It is not only a clinical scale, but also a reliability scale. In fact, the score of this scale is included into the evaluation of the reliability of the test since a high level of anxiety is supposed to alter the

performance of the patient causing unreliable answers.

3. *Sheet* 3: State Trait Anxiety Inventory X2 (STAI-X2); it is the trait anxiety scale of the Italian version of Spielberger's questionnaire (Spielberger et al., 1970; Lazzari & Pancheri, 1980). This is the first totally clinical scale and it investigates the personality trait of anxiety described as a way to face many situations during the lifetime. Thus, the construct under analysis in this scale is supposed to be quite stable and it represents a strong characteristic of the patient's lifestyle.

4. *Sheet* 4: anamnesis questionnaire; it includes 59 wide spectrum items useful in collecting information on several areas of the patient such as family, personal history, relevant life events, health, pain, sleeping habits, eating behavior, etc. This part of the questionnaire represents the main element of horizontal integration of the battery since it is supposed to provide the clinician with a set of accessory information useful to connect the clinical evidences emerged from the other scales.

5. *Sheet* 5: Eysenk Personality Questionnaire Reduced form (EPQ-R; Eysenck & Eysenk, 1975; Sanavio et al., 1986); it is a personality inventory composed by 48 items subdivided into three main scales N (Neuroticism), E (Extroversion) and P (Psychotic). Furthermore, a control scale L (Lie) is introduced and it represents another reliability

index of the battery. Whenever a patient scores high in the L scale he is supposed to simulate trying to provide a too positive image of himself.

6. *Sheet* 6: Psycho-Physiological Questionnaire Reduced form (QPF-R; Sanavio et al., 1986); this questionnaire includes thirty items investigating physiological symptoms. The patient is asked to indicate the frequency of each listed symptom.

7. *Sheet* 7: Fears Inventory (IP-R; Wolpe & Lang, 1964; Sanavio, 1986; Sanavio et al., 1986); it includes 58 items investigating common and specific fears. Seven different indexes can be obtained from this questionnaire: the F score is the general level of fear of the patient, the PH index represents the number of items that extremely scare the patient, the IP1, IP2, IP3, IP4, IP5 indicate the specific score obtained in each area of fears.

8. *Sheet* 8: Depression Questionnaire (QD; Sanavio et al., 1986); it includes 21 items investigating the construct depression. The construct includes: loss of interests, fatiguability, uselessness toughs, etc.

9. *Sheet* 9: Maudsley Obsessional-Compulsive Questionnaire Reduced form (MOCQ-R; Hodgson & Rachman, 1977; Sanavio & Vidotto, 1985); it

is a 21 dichotomous items scale investigating the three main specifications of the Obsessive Compulsive Disorder through three different sub scales, i.e. Checking, Cleaning and Doubting-Ruminating. Furthermore, a general index is obtained by computing the sum of the three sub scales. This questionnaire is the instrument we used to carry out the formalization of FPA.

10. *Sheet* 10: STAI-X1 reduced form; it includes ten out of the twenty items of the second sheet and it is used both to evaluate the state anxiety level at the end of the test and to compare this level with the one measured at the beginning. The score of this scale is used also to compute two further reliability indexes of the battery: the STAI-DIFF representing the difference between the level of state anxiety at the beginning and at the end of the questionnaire; the STAI-ACC indicating the level of coherence in the change between the two administrations of the STAI-X1. The higher the coherence, the higher is the reliability of the assessment.

**Scoring**

The first operation to carry out when evaluating the scores of CBA 2.0 is the computation and examination of the reliability indexes. These indexes are i) the score at STAI-X1, ii) the score at the L scale of EPQ-R, iii) the STAI-DIFF and STAI-ACC indexes, iv) the Repeated items index (IR). Whenever one or more of these indexes is not acceptable the whole battery has to be evaluated with much care knowing that the answers may be not reliable. In this case it is always suggested to check with the patient the reasons of the low level of collaboration to the assessment, or the reason why he's trying to provide a too positive image of himself.

After computing the score of each scale, the clinician has to consider the so called "critical items", i.e. those items that, independently from the score of the scale, represent strong signals to deepen. Some examples of these items are those related to suicide, to homosexuality, to substances abuse, etc.

After the completion of the evaluation of the primary scales the clinician is supposed to hypothesize, when needed, a set of further investigations useful to better understand the critical scores obtained by the patient. These further tools are called "secondary scales" and they investigate specific disorders. Through the secondary scales the battery conducts an attempt of vertical integration of the clinical assessment. For instance, if the patient under analysis presents a critical score at the QD, this datum could be further investigated using the Beck Depression Inventory; if a patient presents a critical score at the MOCQ-R a potential instrument to be administered is the Padua Inventory, and so on.

Another fundamental and innovative characteristic introduced in Italy by CBA 2.0 is the opportunity to have a descriptive computerized report of the score of the patient. This report includes: the analysis of protocol validity and reliability (i.e. the evaluation of the IR index, of the STAI X1, of the STAI-DIFF and STAI-ACC, and the score obtained at the scale L of EPQ-R), the main critical scores obtained by the patient and the areas to further investigate. Furthermore, the answers to the most critical items included in the battery are reported. The very important aspect underlying this descriptive computerized report is an algorithm developed on the basis of a boolean matrix (Boole, 1847/1993, 1854/1976) in which the different patterns of critical scores are related to a specific kind of further investigations. This matrix represents a sort of historical antecedent of FPA. In fact, it delineates a decision tree on the basis of the scores obtained at each single scale of the battery and proposes a decision rule based on previous literature and clinical experience. This algorithm actually performs once what the algorithm of FPA is supposed to perform n-times. This last, but not least, characteristic is probably the main link between CBA 2.0 and FPA.

This in depth exploration of the CBA 2.0 was needed because this battery represents, in our opinion, both the historical antecedent of our approach to psychological assessment and the potential tool to apply FPA. Furthermore, one of its sheets has been used to study all the methodologies presented and explored in the present thesis.

### 1.2.7 Computerized Scoring

In the last decades computers have become one of the most used tools in many fields of human activity. Psychology and psychological assessment have been strongly influenced by the introduction of computers in the clinical activity. Since the beginning of the eighties computers have been used in clinical assessment mostly to perform the scoring of questionnaires in a faster and more accurate way. Together with the improvement of this technology computers have been the more and more used even in the interpretation of the obtained scores (Butcher, Perry, & Atlis, 2000).

As stated above there can be several applications of the computer in psychological assessment: the simplest one is the scoring. The application of computerized techniques to calculate the score of a patient to a specific questionnaire allows for both the reduction of the amount of time needed to score a test (consider, for instance, the time needed to score the MMPI-2), and for a certainly correct calculation of the score. Another way to apply computer in the clinical assessment has been developed more recently and refers to the opportunity to have algorithms performing an initial evaluation of a set of diagnostic indexes in order to provide the clinician with a first overall evaluation of the patients. Finally, we have some electronic version of a number of questionnaires. This last form of computer based questionnaire administration reproduces the paper and pencil version of the questionnaire on a monitor of a computer. The advantages of these versions mostly refer to a more efficient administration of the test (for instance, with patients presenting difficulty with their eyesight the size of the text can be increased), and an immediate scoring and interpretation of the output. On the other hand some limitations are related to the relatively low level of practice with computers of some patients and to the fact that such kind of evaluation could appear dehumanized. This last point could allow, on the contrary, the patient to report even behaviors and information more difficult to collect through an interview (e.g. sexual habits information, relational problems, etc.). A number of studies compared the computerized and the traditional paper and pencil versions of several questionnaires, finding that their reliability and clinical validity are in general overlapping (Finger & Ones, 1999).

### 1.2.8 Adaptive Assessment

In this section we want to introduce one of the key concepts of FPA: adaptivity. One of the main properties of the assessment methodology we are going to introduce is, together with its formal properties, the idea of adaptivity. What we are aimed at building is a computerized adaptive tool, based on a couple of mathematical formal theories able to provide the clinician with the possibility to rely on an efficient and effective vertical integration inference procedure.

Adaptive assessment is based on the fact that the following question posed by the clinician depends on the answer given by the patient to the previous one. In other words, given a first question, the procedure collects the answer of the patient, starts to evaluate the most informative question to pose next, proposes the new question, and so on. Once the procedure completed the evaluation, i.e. no more questions are needed to better specify the problems of the patient, it stops and provides the clinician with an output. Such a procedure is exactly the same underlying any clinical interview. The main difference between such a procedure and a traditional interview is mostly referred to the fact that inferences are made in two very different ways: in the traditional interview inference is left to the clinician who is supposed to take into account all the needed information to proceeded through the interrogation, in the adaptive context inferences are made by an algorithm which considers all the information and goes through the assessment on the basis of logically correct steps. In other words, an adaptive assessment is supposed to be both more efficient and accurate than a traditional assessment. In fact, it is not easy either consider all the information or perform systematically correct inferences. In other words, the introduction of an adaptive computerized assessment should lead to a more accurate and faster psychological assessment without eliminating the critical role of the psychologist who is the only capable of complete the information collected by the computer with all that information coming from the behavioral observation (e.g., non-verbal or para-verbal behaviors).

One could probably be concerned about the reliability of such a methodology. In order to answer to this kind of issues several studies have been conducted and their conclusion was that the computerized adaptive assessment procedures turned out to be at leat as accurate as traditional interviews, but usually more accurate (Grove, Zald, Lebow, Snitz, & Nelson, 2000). In fact, it has been observed that in traditional interviews clinicians tend to to overestimate the weight of some variables, underestimating some others. Furthermore, the clinician cannot have any online information able to warn him of his errors.

Even if the opportuneness to apply adaptive assessment has been proved, the particular frame in which psychological assessment is included has caused a delay in its application. Actually only a very few number of procedures support this kind of feature. Nevertheless, many adaptive assessment tools are present in literature referring to knowledge assessment. The methodology we are going to introduce takes its inspiration from one of the most popular and reliable software used in knowledge assessment, i.e. ALEKS (Assessment and LEarning in Knowledge Spaces). This software was developed to assess mathematical, statistical knowledge and it is now expanding its field of competence to economics and chemistry. The algorithm behind this assessment tool relies on Knowledge Space Theory (Doignon & Falmagne, 1985), one of the two mathematical theories representing the starting point of FPA. In ALEKS given a set of items to be investigated on a specific topic, the output of the assessment (called *knowledge state*) exactly identifies the subset of items that the student under investigation has demonstrated to master. Items are ordered on the basis of their difficulty, thus, this kind of assessment easily informs the teacher on what the student already knows and what he is ready to learn. The system provides the student with a graphical output (called *pie*) displaying, in a very understandable way, how big is his slice of knowledge and how big is the slice to be learned.

In this chapter we were aimed at introducing the state of art about psychological assessment. This introduction was crucial in order to clarify the reason why we tried to create a new methodology able to copy with some of the unresolved issues of the psychological assessment. More specifically we are going to introduce a method that tries to put together the advantages of a questionnaire, of a clinical interview and of the adaptive assessment tools created so far. These advantages will be further improved by some methodological issues that will allow us to collect a greater amount of information out of a set of items administered through an adaptive (i.e. more efficient and accurate in the inference process) algorithm.

In particular we will refer to the possibility to go beyond the mere score of a questionnaire by providing the clinician with a set of specific information about single diagnostic criteria. Up to now, there have been mainly two psychometric theories to be used in evaluating a score: Classical Test Theory (CTT; Novick, 1966), and Item Response Theory (IRT; Rasch, 1980; F. M. Lord, 1980). Both of them take into account measurement-psychometric characteristics mostly related to the score an its reliability. One the one hand CTT, through indexes such as Crombach's alpha or item-total correlation, investigates how each item is a good representation of the construct under investigation, how this construct is well defined and consequently how reliable can be considered the measurement; on the other hand, through indexes as separation index and location, IRT studies how each single item is located along the continuum of difficulty-ability. Furthermore, IRT have been used to construct short-forms of questionnaires able to collect an equivalent amount of information through a lower number of items (Vidotto et al., 2006; Vidotto, Carone, Jones, Salini, & Bertolotti, 2007; Vidotto, RossiFerrario, Bond, & Zotti, 2010; Vidotto, Moroni, et al., 2010). None of these two approaches account for the specific characteristics investigated by each single item, for the specificity of each single response pattern, for the probability to observe an answer that actually does not represent the reality of the patient. We are aimed at introducing an approach (namely FPA) accounting for all these issues and adaptive. In other words the three main characteristics that our approach will have, compared to CTT and IRT are:

1. A higher amount of information provided to the clinician, instead of the mere score;

2. A higher level of reliability and validity of the measurement;

3. Ability to process in a faster way a higher number of information in the vertical integration inference process (computerized adaptive algorithm).

In the next chapter we are going to introduce the two mathematical theories used in FPA with their specific concepts.

# Chapter 2

# Knowledge Space Theory and Formal Concept Analysis: Mathematical Foundations and Potential Overlaps

## 2.1 Introduction

Knowledge Space Theory (KST; Doignon & Falmagne, 1985, 1999) has been applied for the efficient computerized assessment of knowledge and training (Doignon & Falmagne, 1999; Dowling, Hockemeyer, & Ludwig, 1996; Hockemeyer, Held, & Albert, 1998) therefore the main field of application of this theory is education. The computerized procedures built using this theoretical background provide evaluations, in the form of attitudes outlines, showing what the subject is able to do and what he is ready to learn. Two recent papers (Spoto, Stefanutti, & Vidotto, 2010; Spoto et al., 2008) showed how the main concepts of this theory, jointly with the theoretical background of Formal Concept Analysis (FCA; Ganter & Wille, 1999; Wille, 1982), can be employed, starting from the model proposed by Rusch and Wille (1996), in the definition of a methodology for constructing and evaluating the relations between a set of (clinical) items and a set of (clinical) symptoms. More specifically, in this thesis KST has been applied, jointly with FCA, in the definition of a methodology (FPA) for constructing a formal representation of the relation between the items of a given questionnaire. The obtained formal representation can be used to develop an adaptive and efficient tool for psychological assessment. In order to provide practical examples of the proposed methodology we are going to present the application of the method to the items of the Maudsley Obsessive-Compulsive Questionnaire (MOCQ; Hodgson & Rachman, 1977).

The idea behind the methodological proposal we are presenting relies on the opportunity to identify how the main positive aspects of both KST and FCA can be applied to psychological assessment (e.g. clinical assessment,

personality assessment, etc.). Since both theories have been developed and traditionally applied into contexts that are quite different from the one we are going to explore, we previously need to verify whether their main concepts could be translated and adapted to clinical context. In other words we have to reinterpret the main concepts of the theories in light of this field of application and we have to verify whether this "translation" is allowed or not. After this verification we could then test whether the theoretical hypothesized models provide a good representation of the empirical data.
The first of these the controls will be displayed in the present chapter, the second one will be the subject matter of the next two chapters.

## 2.2 Knowledge Space Theory

A *knowledge domain* can be defined as the set $Q$ of all the items that it is possible to investigate about a specific topic. The main authors of KST have focused their attention on topics included in the areas of Mathematics and Statistics, even if they did not exclude the possibility to apply the logical and formal structure of KST to different topics (Doignon & Falmagne, 1999). Some further explanations will clarify both the reason why KST has been mostly applied in those specific contexts, and why it has been used in this study.

Given the knowledge domain $Q$, a *knowledge state* $K \subseteq Q$ represents the subset of $Q$ that a specific subject is able to solve. A *knowledge structure* $\mathcal{K}$ is then defined as a collection of knowledge states including at least the empty set ($\emptyset$) and the total set ($Q$). In the traditional formal notation a knowledge structure is denoted as $(Q, \mathcal{K})$ where $Q$ represents the knowledge domain and $\mathcal{K}$ represents the collection of subsets included in the structure. The knowledge structure is a representation of the implications among the items belonging to $Q$. Using this notation it is possible to identify (i) the field of knowledge that is under consideration, (ii) the specific knowledge presented by a single subject and (iii) the relations that link together the different items of $Q$.

An example may be useful to better understand this last sentence: consider the following knowledge structure defined on a set $Q$ of three items $a$, $b$ and $c$:

$$\mathcal{K} := \{\emptyset, \{a\}, \{b\}, \{a, c\}, \{a, b, c\}\} \tag{2.1}$$

In the knowledge structure $\mathcal{K}$ we can observe that the knowledge domain $Q$ is composed by the three items $a$, $b$ and $c$; the relations among the items in $Q$ determine the admissible knowledge states. In the given example the mastery of item $a$ is a prerequisite for the mastery of item $c$ (i.e. $a$ is a prerequisite for $c$) in fact there is no state in $\mathcal{K}$ containing $c$ and not containing

$a$. In other words any subject failing item $a$ would necessarily fail item $c$ (excluding the possibility of careless errors and lucky guesses).

When the collection $\mathcal{K}$ of knowledge states of a knowledge structure $(Q, \mathcal{K})$ is closed under union (i.e. every union of knowledge states is again a knowledge state included in the structure; formally: $\bigcup \mathcal{F} \in \mathcal{K}$ for all $\mathcal{F} \subseteq \mathcal{K}$), the knowledge structure is then called a *knowledge space*.

An interesting property of a knowledge space is that more than a single set of prerequisites are allowed for an item. This means that the same item can be solved using different solution strategies (Doignon & Falmagne, 1985).

Another fundamental concept of KST is that of a *skill-map* (Albert, Schrepp, & Held, 1992; Doignon & Falmagne, 1999; Falmagne, Koppen, Villano, & Doignon, 1990; Hockemeyer, Conlan, Wade, & Albert, 2003; Lukas & Albert, 1993). The concept of *skill* has been introduced in KST in order allow the theory to go beyond a formal-mathematical interpretation, where the involved cognitive aspects are limited to comprehensive and general notions. The concept of a skill introduces a more detailed cognitive interpretation of the presented elements. The authors of the theory, referring to previous works (Marshall, 1981; Falmagne et al., 1990; Albert et al., 1992; Lukas & Albert, 1993) identified the cognitive concept of skills to describe the set of abilities, methods, strategies, procedures, etc. a subject could follow in order to solve a specific set of items. In our view this concept can be easily adapted and reinterpreted in a clinical context. A skill map is a triple $(Q, S, f)$ where $Q$ is a non-empty set of items, $S$ is a non-empty set of skills and $f$ is a mapping from $Q$ to $2^S \setminus \{\emptyset\}$ (i.e. the powerset of $S$ excluding the empty-set; Doignon & Falmagne, 1999). For any item $q$ in $Q$ the subset $f(q)$ of $S$ represents the set of skills assigned to $q$. In our translation we use the concept of a skill-map and adapt it to the psychological assessment context where the elements of $S$ are interpreted as clinical symptoms rather than abilities needed to solve a specific item.

Starting from this very natural assumption many other authors have used these concepts in several different ways (e.g. Doignon, 1994; Düntsch & Gediga, 1995; Hockemeyer et al., 2003; Spoto, Stefanutti, & Vidotto, 2010). Thus, it is possible to define a *skill map* as a triple $(Q, S, f)$ where $Q$ is a nonempty set of items, $S$ is a nonempty set of skills (in the sense described above) and $f$ is a mapping from the set of items $Q$ to $2^S \setminus \{\emptyset\}$. There are three different models to delineate a knowledge structure through a skill map: the *Disjunctive Model*, the *Conjunctive Model* and the *Competency Model*. We are going to briefly describe the first two models and then go into the details of the third one.

Doignon and Falmagne (1999) demonstrated that a knowledge structure delineated via the disjunctive model by a skill map is a knowledge space, thus, any knowledge space is delineated by at least one skill map. Out of this demonstration it is possible to show that a knowledge state $K \subseteq Q$ is delineated (via the disjunctive model) by $F \subseteq S$ if

$$K = \{q \in Q : f(q) \cap F \neq \emptyset\}. \tag{2.2}$$

Dually it is demonstrated how a knowledge structure delineated via the conjunctive model is the simple closure space of the knowledge space delineated via the disjunctive model.

It follows that a knowledge state $K \subseteq Q$ is delineated (via the conjunctive model) by $F \subseteq S$ if

$$K = \{q \in Q : f(q) \subseteq F\}. \tag{2.3}$$

Among the several different characteristics and properties of these two models we stress that, with a disjunctive model, in order to master an item it is sufficient to have at least one of the required skills, on the other hand, in the conjunctive model, all the skills are necessary in order to achieve a specific item. The disjunctive and conjunctive models are particular cases of a more general model. In fact, the knowledge structures corresponding to these two models are respectively closed under union and intersection. In the present chapter we are going to refer to structures which are neither closed under intersection nor under union. This kind of structures can be delineated by a skill map via the competency model whose central notion is that of a skill multi map. In a skill multi map a collection of subsets of skills, namely $\mu(q)$, is associated to each item $q$ in the domain. Each set of skills $M \subseteq S$ contained in $\mu(q)$ is called a competency and it can be described as a strategy, a method to solve a given item $q$. Note that in this model, in order to master an item, one of its competencies is sufficient, but all the skills contained in that strategy are necessary. Formally, a skill multi map is a triple $(Q, S, \mu)$ where $\mu$ is a mapping from $Q$ to $2^{2^{S \setminus \{\emptyset\}}} \setminus \{\emptyset\}$, associating to each item $q$ a nonempty family $\mu(q)$ of nonempty subsets of $S$ (i.e. a set of competencies). The knowledge states are delineated by a skill multi map $\mu$ via the competency model $F_\mu : 2^S \rightarrow 2^Q$ such that for each $X \subseteq S$

$$F_\mu(X) := \{\exists M \in \mu(q) | M \subseteq X\}. \tag{2.4}$$

From this formulation, again appears that each state $K$ is composed by all those items with at least one of their competencies included in $X$.

All the elements introduced so far describe the "deterministic" case. Here we are interested in testing a structure by evaluating its fit to the data from a "probabilistic" point of view .

All the elements introduced so far are by definition deterministic. Doignon and Falmagne (1999) introduce and define a *probabilistic knowledge structure* (PKS) as a triple $(Q, \mathcal{K}, \pi)$ where: i)$(Q, \mathcal{K})$ is a knowledge structure; ii) $\pi$ is a probability distribution on $\mathcal{K}$. In the *Basic Local Independence Model* (BLIM; Doignon & Falmagne, 1999) the answers to each single item are locally independent given the knowledge states. This model has been applied in a number of different contexts (e.g. Falmagne et al., 1990; Stefanutti, 2006; Spoto, Stefanutti, & Vidotto, 2010)[1]. In the model, starting from the probabilistic structure $(Q, \mathcal{K}, \pi)$ and a response pattern $R \subseteq Q$, a probability distribution can be derived for each $R$ (Falmagne & Doignon, 1988b):

$$p(R) = \sum_{K \in \mathcal{K}} \rho(R, K)\pi(K),$$

where $\rho$ is the response function assigning to each $R$ its conditional probability given a state $K$. The response function $\rho$ satisfies local independence for each $q \in Q$, thus $\rho(R, K)$ is determined on the basis of the two error parameters of the BLIM, i.e. $\alpha$ (careless error) and $\beta$ (lucky guess) by equation (2.5):

$$\rho(R, K) = \left[\prod_{q \in K \setminus R} \alpha_q\right] \left[\prod_{q \in K \cap R} (1 - \alpha_q)\right] \left[\prod_{q \in R \setminus K} \beta_q\right] \left[\prod_{q \in \overline{K \cup R}} (1 - \beta_q)\right]. \tag{2.5}$$

Parameters $\alpha$ and $\beta$ are in general expected to be low. One of the critical elements introduced through the FPA is a methodology to cope with high levels of careless error and lucky guess. This methodology applies to the

[1] For a detailed description of the Basic Local Independence Model refer to Doignon and Falmagne (1999) and to Falmagne and Doignon (2010). In these two books are included both a formal presentation of the model and some applications carried out during the years using BLIM.

most general knowledge structures and allows to consider $\alpha$ and $\beta$ parameters estimates as diagnostic tools of models goodness of fit.

In the next chapters we will present how the concepts introduced so far can be applied to clinical psychology, furthermore we will show the methodological details of the management of the error $\alpha$ and $\beta$ values.

## 2.3 Formal Concept Analysis

The first basic notion of FCA is the *formal context* defined as a triple $(G, M, I)$ where $G$ is a set of *objects*, $M$ is a set of *attributes* and $I$ is a binary relation between the set of objects and the set of attributes. A formal context is usually represented by a Boolean matrix where each row is an object while each column is an attribute. Whenever a value 1 is present in the entry $(g, m)$ it means that the relation $gIm$ holds, in other words it means that the object $g$ has the attribute $m$. Between the objects and the attributes of a formal context a *Galois connection* is defined. For all the sets $A \subseteq G$ and $B \subseteq M$, the following two transformations define the Galois connection:

$$A' := \{m \in M \mid gIm, \forall g \in A\} \tag{2.6}$$

$$B' := \{g \in G \mid gIm, \forall m \in B\} \tag{2.7}$$

In words, $A'$ is the collection of all the attributes that all the objects in $A$ have in common. Dually $B'$ is the collection of all the objects that possess all the attributes in $B$. It is now possible to introduce a fundamental notion of FCA. The pair $(A, B)$ is called a *formal concept* if it satisfies the following two conditions: $A = B'$ and $B = A'$. The so called *extent* $A$ of the formal concept contains exactly those objects of $G$ that have all the attributes in $B$; the so called *intent* $B$ of the formal concept includes exactly those attributes satisfied by all the objects in $A$.

A sub-concept super-concept relation is then defined in the following way:

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \tag{2.8}$$

or equivalently

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow B_1 \supseteq B_2 \tag{2.9}$$

In words, a concept is of a lower level when it has a larger extent (or equivalently a smaller intent).

The concepts of a context form a *complete lattice* (Birkhoff, 1937, 1967) that is called the *concept lattice* of $(G, M, I)$. The intents of a concept lattice are closed under intersection i.e. each intersection of sets of attributes is included in the lattice. Rusch and Wille (1996) show that the collection of the complements of the intents of a formal context is closed under set-union and so it is a knowledge space (Rusch & Wille, 1996). In their article, the authors start from a formal context defined by the set $G$ of subjects (that in this case are treated as formal objects), the set $M$ of items, and the binary relation $gIm$ meaning that the subject $g$ has correctly answered item $m$. In these terms a response pattern can be seen as a set of formal attributes, an intent. By the relation between the intents of a formal context and their complements, the authors derive a so called *knowledge context* having the domain defined by the set of items and the states by the complements of the observed response patterns. Using this methodology it is then possible to construct a knowledge space starting from a formal context.

# Chapter 3

# Knowledge Space Theory, Formal Concept Analysis and Formal Psychological Assessment

## 3.1 Introduction

The matter issue of the present chapter is to introduce a procedure for deriving a knowledge structure from a skill map. This procedure is essentially based on some interesting connections between KST and FCA pointed out by Rusch and Wille (1996).

An example of the joint application of KST and FCA in a clinical context, similar to the one introduced in Spoto et al. (2008), is presented. In this application three knowledge structures representing the starting point for the construction of an adaptive computerized assessment tool are built and empirically tested. Such structures represent the relations between the items of the Maudsley Obsessive-Compulsive Questionnaire (MOCQ; Hodgson & Rachman, 1977) in its reduced version presented by Sanavio and Vidotto (MOCQ-R; 1985), and the diagnostic criteria for Obsessive-Compulsive Disorder (OCD) included in the Diagnostic and Statistical Manual of Mental Disorders IV-TR (DSM-IV-TR; American Psychiatric Association [APA], 1995).

The results of this chapter have to be considered as a mere example of the potential use of the proposed approach and they will represent the starting point for the deepening of FPA introduced in the next chapter. Furthermore, in the specific context, it is shown how the proposed methodology can be used as a tool for assessing a questionnaire's reliability and construct validity.

## 3.2 Methods

### 3.2.1 Structures construction

In this section we describe how existing FCA concept lattice construction algorithms have been applied for deriving the knowledge structure delineated by a given conjunctive *skill map* $(Q, S, f)$. As described in the previous chapter, a knowledge state $K$ is said to be delineated by a subset $X \subseteq S$ of skills via the *conjunctive model* if

$$K = \{q \in Q : f(q) \subseteq X\}. \tag{3.1}$$

The basic idea behind equation (3.1) is that an item can be solved by a set $X$ of skills if all skills $f(q)$ needed by that item are contained in $X$. Therefore, the collection of all items that are solvable by $X$ is the knowledge state $K$. Then the resulting family of all such states is the knowledge structure delineated by $(Q, S, f)$ via the conjunctive model (Doignon & Falmagne, 1999). Two remarks are needed. The first one regards the fact that the "skills" used here are not considered in the sense of "abilities needed to solve a particular sub-set of items", but in the one of "set of diagnostic criteria satisfied by a patient who answers "True" to a particular sub-set of items". Therefore the set $S$ contains diagnostic criteria rather than skills. The second remark is referred to the use of the conjunctive model to depict the relations among sets of attributes and set of items answered. In the conjunctive model it is hypothesized that a subject who responds to an item would present all the attributes investigated by the item.

Another possible approach to this issue is the *disjunctive model* (Doignon & Falmagne, 1999). Using this model the answer to an item implies that the subject displays at least one of the attributes investigated by that item. It is easy to understand how, using this model, a different perspective is taken in looking at the clinician's interpretation of the score obtained by a patient.

It is now important to highlight a fundamental difference between the two models. Using the conjunctive model, affirmative answers are more informative than negative ones. In fact, each affirmative answer indicates that the patient has all the attributes implied by the item. On the other hand, in the disjunctive model, negative answers are more informative than the affirmative ones because they indicate that the patient does not have any of the attributes implied by the item. This consideration has important consequences on the practical side: it seems reasonable to use the conjunctive model when testing a clinical sample and a disjunctive model with a normal population. Although not done in this chapter, it would be interesting to assess the performance of the disjunctive model with the present data-set.

The conjunctive model has been chosen because it represents the clinician's interpretation of the score obtained to a questionnaire in the following sense:

usually a clinician looks at the score obtained by the patient and evaluates whether it is clinically significant or not. In other words, he assumes that the patient displays all the characteristics investigated by the specific scale. Using the conjunctive model it is possible to more deeply investigate the relations between the items on the basis of the attributes that each item satisfies, assuming that whenever a subject answers "True" to an item he displays (or at least he supposes to possess) all the attributes investigated by that item. The structure's construction algorithm rests on the following principles and concepts.

A formal context corresponding to $(Q, S, f)$ can be derived by interpreting $Q$ as the collection of objects, $S$ as the collection of attributes, and by defining a binary relation $R \subseteq Q \times S$ so that, for all pairs $(q, s) \in Q \times S$

$$qRs \quad \Longleftrightarrow \quad s \notin f(q) \tag{3.2}$$

With these basic definitions at hand the triple $(Q, S, R)$ can be regarded as a formal context. According to equation 3.2 the notation $qRs$ should be read as "skill $s$ is not required by item $q$". As an effect of this definition, the intent $q' := \{s \in S \mid qRs\}$ is just the complement of $f(q)$ in $S$ (see in this respect Doignon & Falmagne, 1999, p. 96). The collection $\mathcal{I}$ of all the intents of the concept lattice corresponding to this context could then be obtained by closing under intersection the collection $\{q' : q \in Q\}$ of all object intents. There are a number of different algorithms doing this task in a quite efficient way (Ganter & Wille, 1999, p. 64) and many different programs have been developed to implement these algorithms (Guénoche, 1990; Valtchev, Missaoui, Godin, & Meridji, 2002; Vogt & Wille, 1994). For any subset $X \subseteq S$ the corresponding extent $X'$ is obtained by (3.3). It is easily seen that by (3.2) this collection can also be rewritten as

$$X' = \{q \in Q : f(q) \subseteq S \setminus X\}, \tag{3.3}$$

which, by (3.1), happens to be a knowledge state delineated by $(Q, S, f)$. Then, it is a well-known fact in FCA that $X''$ (i.e. the set of attributes obtained by the application of (3.3) to $X'$) $\in \mathcal{I}$ for all $X \subseteq S$. From this fact it follows that, indeed, $\{Y' : Y \in \mathcal{I}\}$ is the collection of *all* knowledge states delineated by the skill map at issue. To summarize, the construction procedure that has been used in practice was the following one:

1. after defining an appropriate skill map $(Q, S, f)$ for the items at hand, the relation $R$ corresponding to it was obtained by an application of the simple rule (3.2);

2. then the whole concept lattice corresponding to the context $(Q, S, R)$ was produced by means of the software GaLícía (Valtchev et al., 2002);

3. at this point the knowledge structure delineated by $f$ was simply the

collection of all the extents of the generated concept lattice.

The knowledge structure composed by the extents of the concept lattice (i.e. sets of items) is closed under set intersection. It is known (Doignon & Falmagne, 1985, 1999) that in this case each single item has a unique set of prerequisites. In this context, and more generally in the analysis of a questionnaire, it is reasonable to accept this assumption. In fact, each item is defined by a set of attributes and, via the conjunctive model, it is assumed that a person answering positively to an item has all the attributes of that item. Thus, the set of attributes is unique and the minimal response pattern for each item is unique too. This consideration changes when applied to the whole diagnostic process, as it is not realistic to assume that a specific diagnosis is associated to a unique minimal set of symptoms.

### 3.2.2 The Maudseley Obsessive Compulsive Questionnaire

The short form of the Maudsley Obsessive-Compulsive Questionnaire (Sanavio & Vidotto, 1985) included in the assessment battery Cognitive Behavioral Assessment 2.0 (CBA 2.0; Bertolotti, Zotti, Michielin, Vidotto, & Sanavio, 1990) is composed of 21 dichotomous items (True-False) investigating the main characteristics of OCD. The questionnaire is subdivided into three sub-scales investigating three of the main specifications of the disorder.

The first sub-scale is called “Checking”, it is composed of 8 items investigating some habits of controlling and re-controlling many things; for instance item 4 “I must check many times some particular things (e.g. gas or water taps, doors, etc.)”, or item 14 “I usually check things more than once”. The score ranges from 0 to 8 and the clinical cut-off is set at the 95° percentile.

The second sub-scale is called “Cleaning”, it is composed of 9 items investigating the habits of washing, cleaning and sense of contamination; for instance item 3 “When I touch an animal I feel contaminated”, or item 17 “Every morning I spend a lot of time in washing myself completely”. The score ranges from 0 to 9 and the clinical cut-off is set at the 95° percentile.

Finally the third sub-scale is called “Doubting-Ruminating”, it is composed of 4 items investigating the presence of intrusive and disagreeable thoughts, for instance item is 2 “I frequently have disagreeable thoughts and I cannot get rid of them”. The score ranges from 0 to 4 and the clinical cut-off is set at the 95° percentile.

The items of the MOCQ-R have been constructed mostly by referring to the diagnostic criteria for the OCD included in a previous version of the DSM. In order to analyze the items of the MOCQ-R we used the latest version of the manual, the DSM-IV-TR (APA, 1995). The OCD is included in the category of “Anxiety Disorders” and its diagnosis is based on five main criteria. The first criterion has been further sub-dived into two main parts: the first one deals with "Obsessions" while the second one refers to "Compulsions".

In the present chapter the diagnostic criteria of DSM-IV-TR have been used to analyze each single item of the three scales of the MOCQ-R. The result of such analysis is represented by three formal contexts (consisting of three Boolean matrices) where the objects of the context were the items of the sub-scale, while the attributes were the diagnostic criteria of the DSM-IV-TR for the OCD. The underlying relation $(gIm)$ indicates that the item $g$ investigates the criterion $m$. In the next subsection the details of the procedure are described. In particular the method of construction of the structures is described and some descriptive indications about attributes implications revealed by the structures are provided.

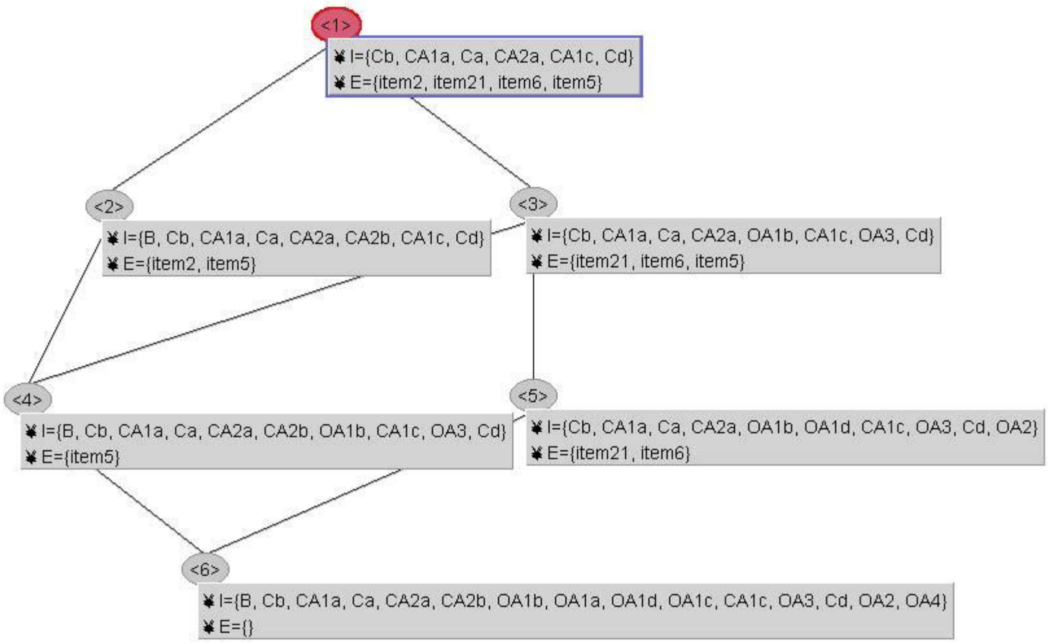
### 3.2.3 The Structures of the MOCQ-R

In Figure 3.1 the concept lattice (i.e. the knowledge structure having sets of items as states) obtained for the sub-scale "Doubting-Ruminating" is presented.

A short explanation of this kind of figure is needed. Figure 3.1 displays the complete lattice obtained for the "Doubting-Ruminating" scale. Each node of the lattice represents a formal concept. The number assigned to each

**Figure 3.1:**

The Concept Lattice obtained for the sub-scale “Doubting-Ruminating”




node is not important (in this thesis). In this chapter we are interested in the collection of objects (items) and attributes (diagnostic criteria) listed in each single gray rectangle. The prerequisite relation among items has to be read bottom-up in the figure. These general rules have to be applied to all the figures included in this thesis.

Going into the details of Figure 3.1, it is interesting to note that there are three different paths from the empty set to the total set of items. More specifically, from the structure it emerges how Item 5 is a prerequisite for Item 2, in fact there is no state including Item 2 and not including Item 5. From the attributes point of view it appears that Item 2 has all the attributes of Item 5 plus some other attributes. In Figure 3.1 it seems that the intent of the set of objects $\{5, 2\}$ (see node 2 of Figure 3.1) contains less elements (i.e. B, Cb, CA1a, Ca, CA2a, CA2b, CA1c, Cd) than the one of the intent of Item 5 (see node 4 of Figure 1; i.e. B, Cb, CA1a, Ca, CA2a, CA2b, CA1c, Cd, OA1b, OA3). In fact the structure of Figure 3.1 is derived by the dual of the formal context having the four items of the sub-scale Doubting-Ruminating as objects, and the diagnostic criteria of DSM-IV-TR as attributes, that is the attributes included in nodes 2 and 4 are not satisfied respectively by the sets of items $\{5, 2\}$ and $\{5\}$.

In Figure 3.2 the concept lattice obtained for the sub-scale "Checking" is presented.

Given this structure it seems that Item 15 ("I follow a very precise order in everything I do") is a prerequisite for most items, and this indicates that the attributes present in it are replicated in many other items of the sub-scale. The remarks on the "Cleaning" structure will be presented later in the "Results" section because some further elements have to be introduced to understand the steps followed to derive the final structure.

### 3.2.4 Testing the Structures

Knowledge structures like the ones presented above are by definition deterministic, they represent a model of possible response patterns of a sample of subjects, but they do not predict the probability of each pattern. As suggested by Doignon and Falmagne (1999) there are two main reasons to introduce probabilities in the model: the first one is that each state should be present with different frequencies in the population; the second one is that the observed response pattern of a subject could not represent his/her real knowledge. From the second reason the opportunity to consider two parameters related to each item follows: the "careless error" (also called "false negative"; $\alpha$) and the "lucky guess" (also called "false positive"; $\beta$) represent respectively the probability that a subject does not solve an item that he is able to solve and the probability of solving an item that he is not able to solve. In other words it is reasonable to introduce conditional probabilities of responses given the knowledge states.

As previously introduced, Doignon and Falmagne define a probabilistic knowledge structure as a triple $(Q, \mathcal{K}, p)$ where: i)$(Q, \mathcal{K})$ is a knowledge structure;

ii) $p$ is a probability distribution on $\mathcal{K}$. In the model at issue, given a state, the responses to the items are locally independent. Thus, starting from the probabilistic knowledge structure $(Q, \mathcal{K}, p)$, given a specific response pattern $R \subseteq Q$ we will define a function $s : (R, K) \mapsto s(R, K)$ assigning to each response pattern its conditional probability given that a subject is in state $K$ (for all states $K \in \mathcal{K}$), the *response function* for the probabilistic knowledge structure. Thus, we obtain for each response pattern a probability distribution:

$$p(R) = \sum_{K \in \mathcal{K}} s(R, K) p(K) \tag{3.4}$$

Since the response function $s$ satisfies local independence for each item $q \in Q$, the conditional probability $s(R, K)$ is determined given the two probabilities $\alpha$ and $\beta$ respectively the careless error and lucky guess related to each item $q$. Formally:

$$s(R, Q) = \left[\prod_{q \in K \setminus R} \alpha_q\right] \left[\prod_{q \in K \cap R} (1 - \alpha_q)\right] \left[\prod_{q \in R \setminus K} \beta_q\right] \left[\prod_{q \in \overline{R \cup K}} (1 - \beta_q)\right] \tag{3.5}$$

Equation 3.5 represents the so called *basic local independence model* (BLIM), which is used in the present chapter.

The parameters of the model have been estimated by a specific version of the Expectation-Maximization Algorithm (Dempster, Laird, & Rubin, 1977) for MatLab, i.e. CEMBLIM. For the description of the algorithm refer to Appendix (7.1).

## 3.3 Results

In order to validate the obtained structures, a data-set provided by a sample of patients from the north eastern part of Italy ($n = 33$; age ranging from 19 to 43 years; 20 males, 13 females) with a diagnosis of OCD has been used. The parameters of the BLIM have been estimated for each of the three structures. The fit of each of the three models has been tested by Pearson's chi-square. It is well known that for large data matrices (as those used in the present study) the asymptotic distribution of $\chi^2$ is not reliable. Therefore a p-value for $\chi^2$ has been obtained by parametric bootstrap(n. of replications= $5,000$).

In the first part of the analysis we tested, for the sub-scale "Cleaning" (9 items), a knowledge structure composed by 80 knowledge states derived by the closure under intersection of the intents of the formal context. The single items' $\alpha$ and $\beta$ parameters seem quite small for almost all items (Table 3.1).

**Table 3.1:**

Estimated parameters $\alpha$ and $\beta$ for each item of the "Cleaning" scale

| Item | 1 | 3 | 8 | 10 | 13 | 16 | 17 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.00 | 0.26 | 0.11 | 0.00 | 0.25 | 0.05 | 0.00 | 0.16 | 0.00 |
| $\beta$ | 0.00 | 0.00 | 0.00 | 0.31 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 |

As previously discussed, in KST $\alpha$ and $\beta$ represent the probability of "careless errors" and "lucky guesses" respectively. It seems more appropriate, in this context, to refer to them as "false negative" and "false positive" in their clinical acceptation.

The results of the bootstrap performed on this model do not support the goodness of fit of the structure ($\chi^2 = 179.98; p = 0.0487$). By the analysis of the content of each single item the deletion of Item 1 appeared to be the best solution to the poor fit. In fact it says "I do not use the public phone because I am afraid of possible contaminations", which seems rather obsolete. The deletion of this item reduces the number of possible states to 40. The new structure is displayed in Figure 3.3

The new structure (i.e. without item 1) has been tested together with the ones obtained for the other two sub-scales. Results show good fit indexes for all three models. Table 3.2 displays the global fit indexes obtained for the three models along with corresponding p-values obtained by parametric bootstrap.

The p-value of the bootstrap performed on the sub-scale "Cleaning" needs some further explanation. The small value observed could be explained by the fact that the number of states derived by closure under intersection of

**Table 3.2:**

The global fit indexes of the three models

| Model | n. of states | *df* | $\chi^2$ | bootstrap $p$ |
|---|---|---|---|---|
| Checking | 18 | 222 | 127.39 | 0.2186 |
| Cleaning | 42 | 198 | 141.65 | 0.1003 |
| Doubting-Ruminating | 6 | 2 | 2.30 | 0.8056 |

the formal context is about 40. This is due to the fact that in this sub-scale the items are rather heterogeneous. They investigate a number of attributes higher than the one investigated by the items of the other sub-scales. The number of states found for the scale “Doubting-Ruminating” is 6 and the number of states found for the sub-scale “Checking” is 19. A larger sample could improve the level of the p-value also because the levels of $\alpha$ and $\beta$ are good especially for the “Cleaning” sub-scale.

Table 3.3 shows the $\alpha$ and $\beta$ parameters obtained for each item in the test.

The overall model fit along with the results displayed in Table 3.3 indicate that the three models quite accurately depict the relations between different items of the MOCQ-R. These relations are well represented by the formal context built using the items of MOCQ-R as objects and the diagnostic criteria of DSM-IV-TR for the OCD as attributes. The two critical parameters,

**Table 3.3:**

Estimated parameters $\alpha$ and $\beta$ for each item of the three sub-scales

| Checking | | | Cleaning | | | Doubting-Ruminating | | |
|---|---|---|---|---|---|---|---|---|
| Item | $\alpha$ | $\beta$ | Item | $\alpha$ | $\beta$ | Item | $\alpha$ | $\beta$ |
| 4 | 0.00 | 0.00 | 3 | 0.27 | 0.07 | 2 | 0.00 | 0.40 |
| 7 | 0.24 | 0.23 | 8 | 0.21 | 0.00 | 5 | 0.06 | 0.00 |
| 9 | 0.06 | 0.13 | 10 | 0.00 | 0.21 | 6 | 0.22 | 0.00 |
| 11 | 0.00 | 0.26 | 13 | 0.23 | 0.27 | 21 | 0.10 | 0.09 |
| 12 | 0.00 | 0.40 | 16 | 0.05 | 0.00 | | | |
| 14 | 0.00 | 0.00 | 17 | 0.00 | 0.00 | | | |
| 15 | 0.16 | 0.00 | 18 | 0.24 | 0.00 | | | |
| 19 | 0.00 | 0.20 | 20 | 0.16 | 0.00 | | | |

i.e. the high values of "false positive" estimates for items 2 and 12 can be explained in two different but not necessarily exclusive ways: the first one refers to the small number of subjects composing the sample. The second one relates to the fact that the items are clinical, thus a sort of misinterpretation of the meaning of the items is possible. This fact can be better understood by looking at the text of the items. Item 2 says "I frequently have disagreeable thoughts and I cannot get rid of them". This item is composed by two separate sentences: the first one is "I frequently have disagreeable thoughts", the second one is "I cannot get rid of them". In the conjunctive model a subject is supposed to provide a positive answer when he has all the elements required. In this case the subject may have answered "True" even if he believed that only one of the two sentences was true. Item 12 says "One of the greatest problems of mine is the repeated check of things". In this case some problems may arise from the interpretation of either "greatest" or "repeated". Anyway, since the analyzed questionnaire is clinical, it seems reasonable to expect higher $\alpha$ and $\beta$ values than in the classical field of application of KST. In fact a subject could intentionally fake the answer. Furthermore the subject's answer could be affected by his poor introspection capabilities.

## 3.4 Discussion

In this chapter we provided some interesting results both from the clinical and the methodological point of view.

From the methodological point of view, the proposed analysis can be used as a tool to validate the construct and content validity of a given questionnaire. In the presented example the validity of MOCQ-R has been assessed on the basis of the presence-absence of the criteria of the DSM-IV-TR which gives a theoretical-clinical interpretation of the questionnaire. The joint application of FCA and KST can evaluate whether a questionnaire actually measures the underlying construct (in this case the Obsessive-Compulsive Disorder). In the example, the underlying construct was the set of diagnostic criteria for the OCD. In this perspective the validity analysis rests on verifying the relation between the items and the criteria (i.e. the attributes of the formal context).

In typical applications a clinician uses the questionnaire by considering the score obtained by the patient. Since the underlying construct is multidimensional, the mere score of a patient is a dramatic reduction of the potential information provided by the test. By the proposed approach the information collected by the questionnaire can be used to point out differences between patients that otherwise would be hidden by the simple score. Indeed, from the clinical perspective, the proposed methodology could be regarded as an in depth evaluation of the construct investigated by the questionnaire. The response patterns corresponding to clinically significant scores (i.e. $> 95°$ percentile) could point to different collections of diagnostic symptoms, and these differences are not captured by the simple score. For instance it is possible to note that the two collections of items $\{3, 8, 10, 13, 16, 18, 20\}$ and $\{3, 8, 13, 16, 17, 18, 20\}$ (corresponding to nodes 2 and 4 of Figure 3.3 respectively) are equally scored 7, but they correspond to different collections of attributes. This means that patients obtaining the same score might have different sets of symptoms.

In the presented case, the relation between the items of the MOCQ-R and the criteria of the DSM-IV-TR has been constructed assuming a conjunctive model. As discussed in the "Structure Construction" section this assumption seems reasonable when a sample of clinical patients is assessed. There are other situations in which the disjunctive model could be more appropriate, for instance when reasoning in terms of a whole diagnostic process in a wider spectrum situation (i.e. a complete battery for clinical assessment).

Considering the practical application of the approach proposed, in a clinical setting the relations found between items could be used to calibrate an algorithm for the adaptive and efficient evaluation of patients. KST has been developed with the aim to construct an efficient tool for assessing knowledge. Some computerized algorithms have been developed with this aim (Falmagne & Doignon, 1988a, 1988b; Dowling et al., 1996; Hockemeyer et al., 1998). These algorithms can be easily adapted to the clinical context for which the structures of Figures 3.1, 3.2 and 3.3 have been developed.

A final remark concerns the $\alpha$ and $\beta$ probabilities of the items. Since they are interpreted as false positives and false negatives, these probabilities are clearly expected to be small. Large values of these parameters could point to bad specification of the model or to bad wording of the items. Therefore these parameters can be used as a diagnostic tool for improving the model or the items. For instance, when a large value of the parameter $\alpha$ is observed for a given item, there might be one or more attributes both necessary to positively answer to a specific item and not included in the model for that item. In other words even if in a given model the attributes referred to a specific item $i$ are for instance $\{a, b, c\}$ a large value of $\alpha$ indicates that one or more further attributes have to be displayed by a subject to answer $i$. There are two main ways to cope with this problem: the first one is to reformulate the model with respect to $i$, including the necessary attributes; the second one is to reword the specific item so that it is in accord with the model. Working with standardized questionnaires the first solution might be preferred, but when a questionnaire is under construction the second approach could be the most interesting one.

Some different issues have to be considered when coping with a large value of $\beta$. There are two main explanations of this. The first one is that in the model one or more attributes have been wrongly considered necessary to answer "True" to a specific item. In this case it is possible to eliminate from the set of the attributes of that item all those attributes that are not necessary for it; otherwise an alternative could be the rewording of the item. The second explanation is that the conjunctive model is not a good representation of the relation between items and attributes. In this case the possibility to apply a disjunctive model can be considered.

Furthermore, in the next chapter we are going to in depth investigate some methodological issues to formally manage the $\alpha$ and $\beta$ values in particular knowledge structures. In the same chapter we will briefly introduce and explore the opportunity to introduce more than one latent class in the model testing, and display the results of the investigation of the probability of each single attribute in each single latent class.

In conclusion, this chapter has highlighted how the proposed application to a clinical context of KST and FCA seems to be a useful and fruitful research and clinical perspective. It has to be further investigated (using larger samples and different clinical disorders) whether the proposed methodology would be able to provide the clinician with the opportunity to perform a more in depth analysis of patients' responses. This opportunity will allow the clinician to do personalized diagnosis able to pinpoint subject specific characteristics. In the long term this methodology represent one of the possible paths to allow the construction of an adaptive, efficient and effective tool for psychological diagnosis. This instrument will be very similar to the instruments already built for knowledge assessment.

**Figure 3.2:**

The Concept Lattice obtained for the sub-scale "Checking"

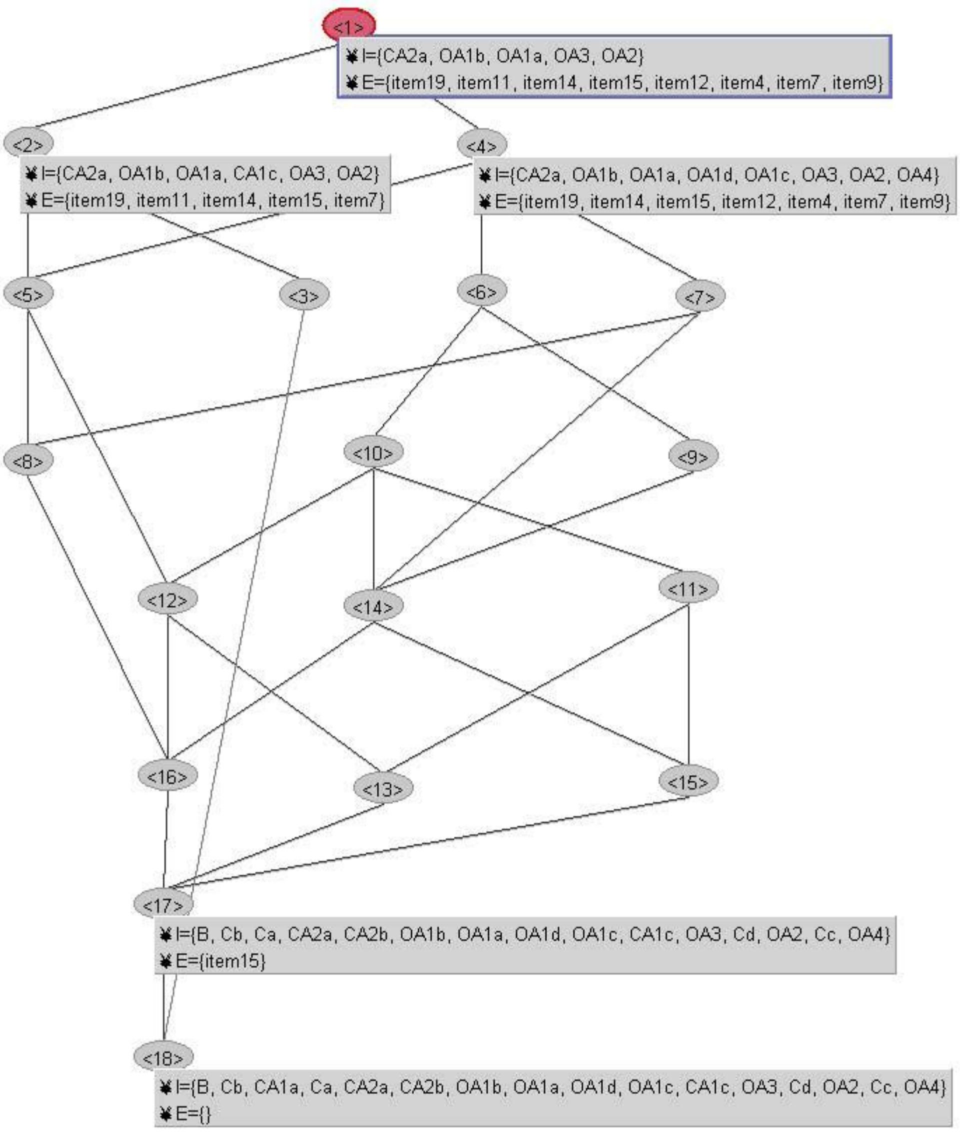
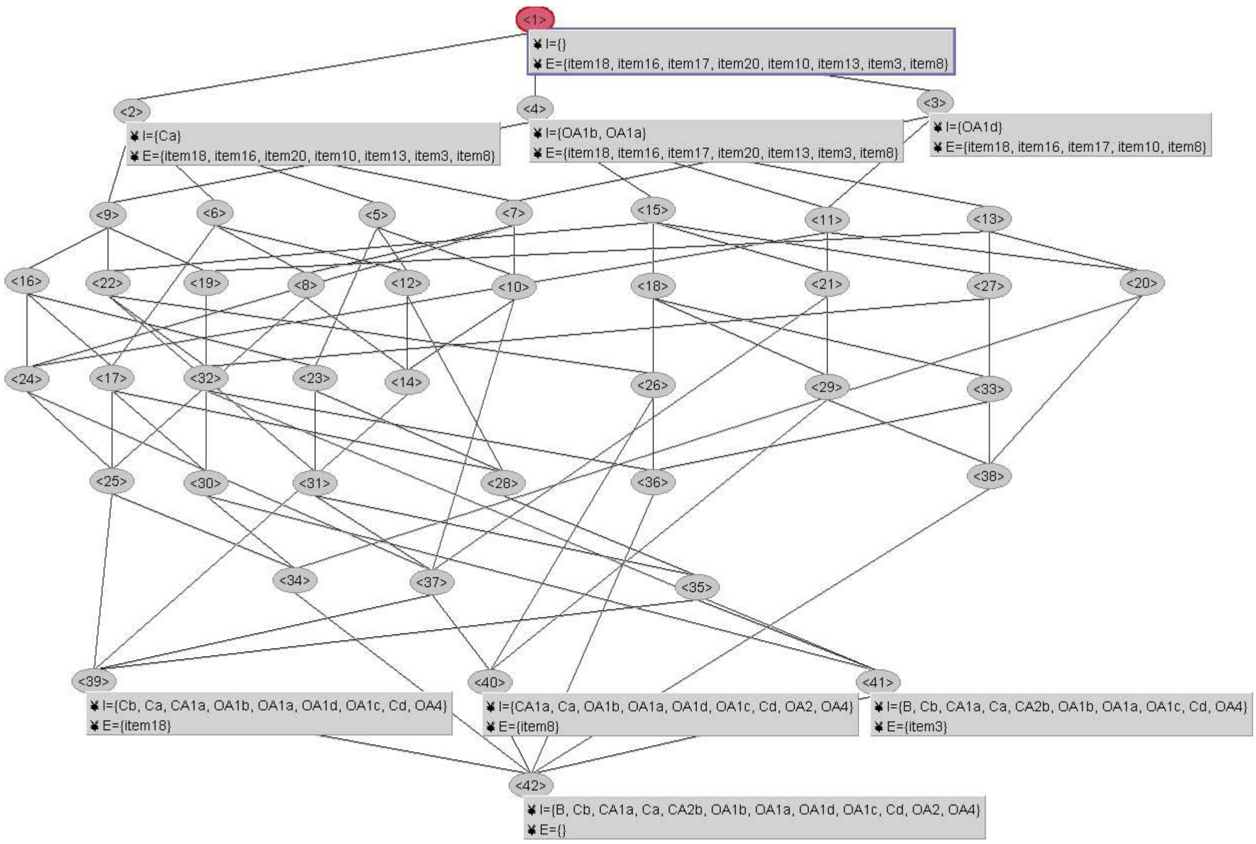


**Figure 3.3:**

The Concept Lattice obtained for the sub-scale “Cleaning” excluding Item 1



# Chapter 4

# Deepening Formal Psychological Assessment

## 4.1 Introduction

The subject matter of the present chapter is twofold: one the one hand we would like to introduce a methodological procedure aimed at managing the careless error ($\alpha$) and the lucky guess ($\beta$) parameters in a knowledge structure; on the other hand we will discuss a development of the results of the previous chapter based on the possibility to both evaluate each single attribute probability and introduce more than one latent class in the model under analysis. More specifically it will be shown how the innovative methodological approach can be applied to the knowledge structure derived from the application of a skill multi map, whose details are given in a moment.

Thus, in this chapter, after presenting an extension of the BLIM to skill multi maps, we will introduce a methodology for deflating the error rates ($\alpha$ and $\beta$) of the model. We will also introduce a suggestion for using these error parameters as diagnostic tools for the goodness of fit of the model under analysis. Both methodological and theoretical fallouts of this approach will be explored. Furthermore, we will present an example of how this methodology can be practically used. An application carried out on a set of clinical data will be shown, referring to a questionnaire which investigates the Obsessive-Compulsive Disorder (OCD), as described by the DSM-IV-TR (American Psychiatric Association [APA], 1995). The questionnaire is again the Maudsley Obsessional-Compulsive Questionnaire (MOCQ; Hodgson & Rachman, 1977) in its reduced form (MOCQ-R; Sanavio & Vidotto, 1985), included in the Cognitive Behavioural Assessment 2.0 (CBA 2.0; Bertolotti et al., 1990). After presenting such application, we will explore the results obtained through the introduction of two latent classes and we will describe the particular dimensionality obtained, even referring to the estimate probabilities of each single attribute of the model into each single latent class. The picture that will emerge will be deeply discussed in order to generalize the observed results to conditions in which more classes are introduced.

## 4.2 Extending the BLIM to skill multi maps

It has been previously stressed how KST, before the introduction of the concept of skill, was a behavioral theory evaluating the relations among items through the observed responses. The same remark can be addressed to the BLIM that is focused on the behavioral part of the response process and thus it can be used to test a probabilistic knowledge structure. Nonetheless, the extension of this probabilistic model to the skill map and skill multi map levels is quite simple.

Some models aimed at describing the relation (in a probabilistic sense) between items and skills have already been reported in literature. Among these we will focus on the DINA (Deterministic Input Noisy And) model for conjunctive skill maps (Junker & Sijtsma, 2001; Tatsuoka, 1985). More recently the DINA model has been extended to multiple solution strategies (DeLaTorre & Douglas, 2008). Such extension turns out to be very rigid since all the items have to present the same number of strategies and the definition of more than two strategies makes the application of the model very complicated.

The theoretical foundations behind the skill maps and multi maps can be fruitfully applied within the PKS approach in order to obtain a model which takes into account in a very flexible way the presence of one or more strategies for one or more items. In our application we will estimate the usage probability of each strategy (either related or not with its difficulty). Furthermore, the proposed approach only needs the construction of one skill multi map. From a formal point of view we will have that, given a response pattern $R \subseteq Q$, and a set of skills $C \in \mathcal{C}$, where $\mathcal{C}$ is a competence structure, the probability to observe a given pattern $R$ under conditional independence of the responses to the items given the competence states, is:

$$P(R \mid C) = \prod_{q \in Q} \left[\alpha_q^{1-w(R,q)}(1-\alpha_q)^{w(R,q)}\right]^{y(C,q)} \times \left[\beta_q^{w(R,q)}(1-\beta_q)^{1-w(R,q)}\right]^{1-y(C,q)}. \tag{4.1}$$

where

$$w(R,q) = \begin{cases} 1 & \text{if } q \in R \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

and

$$y(C,q) = \begin{cases} 1 & \text{if } X \in C \text{ for some } X \in \mu(q) \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

This means that $y(C,q)$ has value 1 when $C$ solves $q$ and 0 otherwise. This extension of the model contains the following parameters: $\alpha$ (careless error), $\beta$ (lucky guess) and a probability $P(C)$ for each competence state. In our application independence among skills will be assumed. This implies that $\mathcal{C}$ is the powerset on $S$, and that, for $C \in \mathcal{C}$, the probability $P(C)$

decomposes as follows:

$$P(C) = \prod_{s \in C} \pi_s \prod_{s \in S \setminus C} (1 - \pi_s), \tag{4.4}$$

where $\pi_s$ is a parameter representing the probability of skill $s$. This assumption is not a necessary condition for the model to be applied.

## 4.3 Deflating error rates in knowledge structures graded in an item

When fitting the BLIM model (or some specific restriction of it) to the data, standard goodness-of-fit statistics like the Chi-square or the likelihood ratio, only partially inform about the correctness of the fitted model. Irrespectively of how good the fit is, a too high value of the $\alpha$ or $\beta$ parameters (say, higher than .5) might be the symptom of a misspecification of the fitted model. As argued by Stefanutti and Robusto (2009), this happens because such type of parameters tend to inflate for misspecified knowledge structures, with the effect of improving their fit to the data.

There are certain types of knowledge structures for which at least one of the two error rates $\alpha$ or $\beta$ of some items can be arbitrarily inflated, deflated or brought down to zero, while preserving the initial goodness of fit of the model. Such types of knowledge structures satisfy a certain kind of "gradation" with respect to some items. They are the subject matter of this section.

### 4.3.1 Equivalent probabilistic representations

In order to study the class of knowledge structures mentioned above, some notation is needed, which is primarily focused on establishing, in a well-defined sense, when two probabilistic knowledge structures are equivalent. Given an arbitrary knowledge structure $(Q, \mathcal{K})$, we indicate with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ the careless error and lucky guess parameter vectors respectively, and with $\pi$ a probability distribution on the knowledge states in $\mathcal{K}$. Moreover call $\kappa := (Q, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \pi)$ a *probabilistic representation* of $(Q, \mathcal{K})$. Within this representation, the probabilities $f_\kappa(R)$ of the response patterns $R \subseteq Q$ are

$$f_\kappa(R) = \sum_{K \in \mathcal{K}} \rho(R, K)\pi(K)$$

where $\rho_\kappa(R, K)$ is the conditional probability of pattern $R$, given state $K$, under the local independence assumption of the BLIM (equation 2.5).

**Definition 1.** Two distinct probabilistic representations $\kappa$ and $\kappa'$ of the same knowledge structure $(Q, \mathcal{K})$ are said to be *equivalent* if $f_\kappa(R) = f_{\kappa'}(R)$ for each response pattern $R \subseteq Q$.

### 4.3.2 Knowledge structures backward-graded in an item

The first class of knowledge structures considered in this chapter, have a probabilistic representation in which, under certain conditions, the careless error of some item is exactly zero.

**Definition 2.** For $q \in Q$, the knowledge structure $(Q, \mathcal{K})$ is called *backward-graded* in $q$ if $K \setminus \{q\}$ is in $\mathcal{K}$ for every state $K \in \mathcal{K}$.

To give an example, the knowledge structure

$$\mathcal{L} := \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{1,2,3,4\}\}$$

is backward-graded in item 3. In fact, by removing item 3 from any state in $\mathcal{L}$ we obtain a subset which is still a member of $\mathcal{L}$. However $\mathcal{L}$ is not backward-graded in item 4. For instance, if this item is removed from state $\{1,3,4\}$ a subset is obtained which is not a member of $\mathcal{L}$.

In proving the main result of this section the following two collections will be often used. Given a knowledge structure $(Q, \mathcal{K})$, let

$$\mathcal{K}_q := \{K \in \mathcal{K} : q \in K\}$$

be the set of all states in $\mathcal{K}$ containing item $q \in Q$, and

$$\mathcal{K}_{-q} = \{K \setminus \{q\} : K \in \mathcal{K}_q\},$$

be the collection of all subsets that can be obtained by removing $q$ from the states in $\mathcal{K}_q$. For notational convenience, for $K \in \mathcal{K}_{-q}$ the shortcut $K_q$ will be used to denote the disjoint union $K \cup \{q\}$.

It is worth noticing that the function $a : \mathcal{K}_q \to \mathcal{K}_{-q}$ such that $a(K) = K \setminus \{q\}$ is a bijection and that, if $(Q, \mathcal{K})$ is backward-graded in $q$ then $\mathcal{K}_{-q} \subseteq \mathcal{K}$.

**Theorem 1.** Let knowledge structure $(Q, \mathcal{K})$ be backward-graded in $q \in Q$, and let $\kappa = (Q, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \pi)$ be a probabilistic representation of it. Define a new representation $\kappa' = (Q, \mathcal{K}, \boldsymbol{\alpha}', \boldsymbol{\beta}', \pi')$ such that, for each $i \in Q$:

$$\alpha'_i = \begin{cases} \alpha_i & \text{if } i \neq q \\ \\ 0 & \text{if } i = q \end{cases}$$

and $\beta'_i = \beta_i$. Moreover, for each $K \in \mathcal{K}$

$$\pi'(K) = \begin{cases} \pi(K) - \delta_q \pi(K) & \text{if } K \in \mathcal{K}_q \\ \\ \pi(K) + \delta_q \pi(K_q) & \text{if } K \in \mathcal{K}_{-q} \\ \\ \pi(K) & \text{elsewhere} \end{cases} \tag{4.5}$$

where

$$\delta_q = \frac{\alpha_q}{1 - \beta_q}.$$

For $\alpha_q < 1 - \beta_q$, the two representations $\kappa$ and $\kappa'$ are equivalent.

*Proof.* Define the two functions $g_\kappa : 2^Q \to [0, 1]$ and $h_\kappa : 2^Q \to [0, 1]$ in the following way: for $R \subseteq Q$

$$g_\kappa(R) = \sum_{K \in \mathcal{K}_{-q}} \rho_\kappa(R, K)\pi(K) + \sum_{K \in \mathcal{K}_q} \rho_\kappa(R, K)\pi(K)$$

and

$$h_\kappa(R) = \sum_{K \in \mathcal{K} \setminus (\mathcal{K}_q \cup \mathcal{K}_{-q})} \rho_\kappa(R, K)\pi(K).$$

Then, from $\mathcal{K}_q \cup \mathcal{K}_{-q} \subseteq \mathcal{K}$, it follows that $f_\kappa(R) = g_\kappa(R) + h_\kappa(R)$ for each $R \subseteq Q$. Moreover, given that $\pi'(K) = \pi(K)$ for each $K \in \mathcal{K} \setminus (\mathcal{K}_q \cup \mathcal{K}_{-q})$, it is also clear that $h_\kappa(R) = h_{\kappa'}(R)$ for each $R \subseteq Q$. Hence the equality $g_\kappa(R) = g_{\kappa'}(R)$ is a necessary and sufficient condition for $f_\kappa(R) = f_{\kappa'}(R)$. In turn, the condition $g_\kappa(R) = g_{\kappa'}(R)$ holds true if the equality

$$\rho_\kappa(R, K_q)\pi(K_q) + \rho_\kappa(R, K)\pi(K) = \rho_{\kappa'}(R, K_q)\pi'(K_q) + \rho_{\kappa'}(R, K)\pi'(K). \tag{4.6}$$

holds for each state $K \in \mathcal{K}_{-q}$.

Consider at first any response pattern $R \subseteq Q$ not containing item $q$, and a state $K \in \mathcal{K}_{-q}$. Since $q \in K_q \setminus R$, for (2.5) we can write $\rho_\kappa(R, K_q) = C\alpha_q$, where the quantity $C \in [0, 1]$ depends on the parameters $\alpha$ and $\beta$ of the items in $Q \setminus \{q\}$. Moreover, for the fact that $q \in \overline{K \cup R}$, still for (2.5), we have $\rho_\kappa(R, K) = C(1 - \beta_q)$. For similar reasons one also has $\rho_{\kappa'}(R, K_q) = C\alpha'_q$ and $\rho_{\kappa'}(R, K) = C(1 - \beta'_q)$. Therefore, (4.6) can be rewritten as:

$$C\alpha_q\pi(K_q) + C(1 - \beta_q)\pi(K) = C\alpha'_q\pi'(K_q) + C(1 - \beta_q)\pi'(K)$$

since the constant $C$ vanishes and $\alpha'_q = 0$, the equality above simplifies to

$$\alpha_q\pi(K_q) + (1 - \beta_q)\pi(K) = (1 - \beta'_q)\pi'(K),$$

and, solving for $\pi'(K)$,

$$\pi'(K) = \frac{\alpha_q}{1-\beta_q}\pi(K_q) + \pi(K). \tag{4.7}$$

For (4.5) we know that this equality holds true, and this shows that when $R$ does not contain $q$, equation (4.6) holds true.

Consider now the response pattern $R_q := R \cup \{q\}$. Since $q \in K_q \cap R_q$ we have $\rho_\kappa(R_q, K_q) = C(1-\alpha_q)$ for some $C \in [0,1]$, and from $q \in R_q \setminus K$, it follows that $\rho_\kappa(R_q, K) = C\beta_q$. Moreover: $\rho_{\kappa'}(R_q, K_q) = C(1-\alpha'_q)$ and $\rho_{\kappa'}(R_q, K) = C\beta'_q$. Substituting in (4.6) and simplifying, one obtains:

$$(1-\alpha_q)\pi(K_q) + \beta_q\pi(K) = (1-\alpha'_q)\pi'(K_q) + \beta_q\pi'(K)$$

and, because $\alpha'_q = 0$,

$$(1-\alpha_q)\pi(K_q) + \beta_q\pi(K) = \pi'(K_q) + \beta_q\pi'(K).$$

Now, for (4.7), we know that $\pi'(K) = \pi(K) + \delta_q\pi(K_q)$, thus

$$(1-\alpha_q)\pi(K_q) + \beta_q\pi(K) = \pi'(K_q) + \beta_q[\pi(K) + \delta_q\pi(K_q)]$$

that, after some algebra, yields

$$\pi'(K_q) = \pi(K_q) - \delta_q\pi(K_q),$$

which is known to be true for (4.5).

We have considered all the states included in $\mathcal{K} \cup \mathcal{K}_{-q}$. We can now analyze the states included in $\mathcal{K} \setminus (\mathcal{K} \cup \mathcal{K}_{-q})$. Notice that none of them include $q$. For all these states we have that $\rho_\kappa(R, K) = \rho_{\kappa'}(R, K) = C(1 - \beta_q)$ and $\rho_\kappa(R_q, K) = \rho_{\kappa'}(R_q, K) = C\beta_q$. It is evident that these probabilities never depend on neither $\alpha_q$ nor $\alpha'_q$. Thus, the satisfaction of the equality (4.6) follows.

Having shown that (4.6) holds true for each pattern $R \subseteq Q$ and each state $K \in \mathcal{K}$ we conclude that $\kappa$ and $\kappa'$ are equivalent probabilistic representations of $(Q, \mathcal{K})$. $\square$

Theorem 1 sheds light on some aspects of a backward-graded knowledge structure that are worth considering. In the first place its probabilistic representation is not unique and, hence, it is not identifiable. However there exists a particular representation $\kappa'$ where the parameter $\alpha'_q$ vanishes. This representation is superior to each of its alternatives in the sense that: (i) it is more parsimonious (it has one parameter less than the other representations) and (ii) the parameter that vanishes is an error probability, something which is desired to be as small as possible.

Theorem 1 is also at the basis of a procedure that can be applied with the aim of zeroing the $\alpha$ parameter of a specific item in an arbitrary (i.e. not necessarily backward-graded) knowledge structure.

Let $q \in Q$ be an item, $\mathcal{K}$ be a knowledge structure on $Q$, and $\kappa = (Q, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \pi)$

be its probabilistic representation. Define a new knowledge structure in the following way:

$$\mathcal{K}' := \mathcal{K} \cup \{K \setminus \{q\} : K \in \mathcal{K}_q\}. \tag{4.8}$$

The structure obtained this way happens to be backward-graded in $q$. Define moreover the following provisional probabilistic representation for $(Q, \mathcal{K}')$: $\alpha_i^* = \alpha_i$ and $\beta_i^* = \beta_i$ for each item $i \in Q$, and

$$\pi^*(K) = \begin{cases} \pi(K) & \text{if } K \in \mathcal{K} \\ 0 & \text{if } K \notin \mathcal{K} \end{cases}$$

for each $K \in \mathcal{K}'$. We obtain, in this way an intermediate probabilistic representation for $\mathcal{K}'$ that we call $\kappa^* = (Q, \mathcal{K}', \alpha^*, \beta^*, \pi^*)$. It is easy to see that $\kappa$ and $\kappa^*$ are equivalent, since $\pi^*(K) = 0$ for every state $K \in \mathcal{K}' \setminus \mathcal{K}$.

At this point, since $\mathcal{K}'$ is backward-graded in $q$, Theorem 1 can be applied to $\kappa^*$, leading to the construction of a third representation $\kappa' = (Q, \mathcal{K}', \alpha', \beta', \pi')$, equivalent to $\kappa^*$ (and thus also to $\kappa$), where $\alpha'_q = 0$.

### 4.3.3 Knowledge structures forward-graded in an item

The second class of knowledge structures considered in the present chapter, have a probabilistic representation where, under certain conditions, the lucky guess of some item is exactly zero.

**Definition 3.** For $q \in Q$, the knowledge structure $(Q, \mathcal{K})$ is called *forward-graded* in $q$ if $K \cup \{q\} \in \mathcal{K}$ for every state $K \in \mathcal{K}$.

As an example, consider the knowledge structure

$$\mathcal{M} = \{\emptyset, \{1\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}, \{1,2,3,4\}\}.$$

This knowledge structure is forward-graded in item 1. In fact $K \cup \{1\}$ is in $\mathcal{M}$ for all states $K \in \mathcal{M}$. However it is not forward-graded in item 2 since $\emptyset \cup \{2\}$ is not a state. It is also worth observing that $\mathcal{M}$ is not backward-graded in item 1. In fact, if 1 is removed from $\{1,2,3,4\}$, the result is not a state. Therefore Definitions 2 and 3 are not equivalent.

Nonetheless there is a strong connection between the two families. Recalling that the dual $\mathcal{K}^\partial$ of a knowledge structure $\mathcal{K}$ is the collection of the complements of the states in $\mathcal{K}$:

$$\mathcal{K}^\partial := \{Q \setminus K : K \in \mathcal{K}\},$$

we have the following

**Theorem 2.** A knowledge structure $\mathcal{K}$ on a set $Q$ is forward-graded on an item $q \in Q$ if and only if its dual is backward-graded in $q$.

*Proof.* Let $\mathcal{K}$ be forward-graded in $q$ and let $K$ be a state in $\mathcal{K}$. Then (i) $K \cup \{q\}$ is in $\mathcal{K}$ and (ii) $\bar{K} := Q \setminus K$ is in $\mathcal{K}^\partial$. From (i) it also holds that $Q \setminus (K \cup \{q\})$ is in $\mathcal{K}^\partial$. But $Q \setminus (K \cup \{q\}) = \bar{K} \setminus \{q\}$. Thus $\bar{K} \setminus \{q\} \in \mathcal{K}^\partial$. $\square$

The following two collections will be used in proving the main result in this section:

$$\bar{\mathcal{K}}_q := \{K \in \mathcal{K} : q \notin K\}$$
$$\bar{\mathcal{K}}_{+q} := \{K \cup \{q\} : K \in \bar{\mathcal{K}}_q\}.$$

We also notice that the function $b : \bar{\mathcal{K}}_q \to \bar{\mathcal{K}}_{+q}$ such that $b(K) = K \cup \{q\}$ is a bijection and that, if $\mathcal{K}$ is forward-graded in $q$, then $\bar{\mathcal{K}}_{+q} \subseteq \mathcal{K}$.

**Theorem 3.** Let knowledge structure $(Q, \mathcal{K})$ be forward-graded in $q \in Q$, and let $\kappa = (Q, \mathcal{K}, \alpha, \beta, \pi)$ be a probabilistic representation for it. Define a new representation $\kappa' = (Q, \mathcal{K}, \alpha', \beta', \pi')$ such that, for each item $i \in Q$: $\alpha'_i = \alpha_i$,

$$\beta'_i = \begin{cases} \beta_i & \text{if } i \neq q \\ 0 & \text{if } i = q \end{cases}$$

Furthermore, for each $K \in \mathcal{K}$,

$$\pi'(K) = \begin{cases} \pi(K_q) + \gamma_q \pi(K) & \text{if } K \in \bar{\mathcal{K}}_{+q} \\ \pi(K) - \gamma_q \pi(K) & \text{if } K \in \bar{\mathcal{K}}_q \\ \pi(K) & \text{elsewhere} \end{cases} \tag{4.9}$$

where

$$\gamma_q = \frac{\beta_q}{1 - \alpha_q}.$$

Under the condition $\beta_q < 1 - \alpha_q$, the two representations $\kappa$ and $\kappa'$ are equivalent.

*Proof.* The proof to this theorem parallels that of Theorem 1 and it is given here for completeness. The two functions of reference are, in this case:

$$g_\kappa(R) = \sum_{K \in \bar{\mathcal{K}}_q} \rho_\kappa(R, K)\pi(K) + \sum_{K \in \bar{\mathcal{K}}_{+q}} \rho_\kappa(R, K)\pi(K)$$

and

$$h_\kappa(R) = \sum_{K \in \mathcal{K} \setminus (\bar{\mathcal{K}}_{+q} \cup \bar{\mathcal{K}}_q)} \rho_\kappa(R, K)\pi(K).$$

Then the condition $g_\kappa(R) = g_{\kappa'}(R)$ holds true if equation (4.6) holds for all states $K \in \bar{\mathcal{K}}_q$.

We first consider $R_q$ for some pattern $R \subseteq Q \setminus \{q\}$, and a state $K_q \in \bar{\mathcal{K}}_{+q}$. Then equation (4.6) can be rewritten as:

$$(1 - \alpha_q)\pi(K_q) + \beta_q \pi(K) = (1 - \alpha'_q)\pi'(K_q) + \beta'_q \pi'(K).$$

Being $\alpha'_i = \alpha_i$ and $\beta'_q = 0$, the previous equation simplifies to

$$(1 - \alpha_q)\pi(K_q) + \beta_q \pi(K) = (1 - \alpha_q)\pi'(K_q),$$

and solving for $\pi'(K_q)$,

$$\pi'(K_q) = \pi(K_q) + \gamma_q \pi(K).$$

Consider now the response pattern $R$ (which does not include $q$) and a knowledge state $K \in \bar{\mathcal{K}}_q$. Then (4.6) takes on the form:

$$\alpha_q \pi(K_q) + (1 - \beta_q)\pi(K) = \alpha'_q \pi'(K_q) + (1 - \beta'_q)\pi'(K)$$

Since $\alpha'_i = \alpha_i$, $\beta'_q = 0$ and $\pi'(K_q) = \pi(K_q) + \gamma_q \pi(K)$, after appropriate substitutions one obtains that

$$\alpha_q \pi(K_q) + (1 - \beta_q)\pi(K) = \alpha_q[\pi(K_q) + \gamma_q \pi(K)] + \pi'(K).$$

Solving for $\pi'(K)$, after some algebra one obtains that $\pi'(K) = \pi(K) - \gamma_q \pi(K)$.

We have considered all the states included in $\overline{\mathcal{K}} \cup \overline{\mathcal{K}}_{+q}$. We can now analyze the states included in $\mathcal{K} \setminus (\overline{\mathcal{K}} \cup \overline{\mathcal{K}}_{+q})$. Notice that all of them include $q$. For all these states we have that $\rho_\kappa(R, K_q) = \rho_{\kappa'}(R, K_q) = C\alpha_q$ and $\rho_\kappa(R_q, K_q) = \rho_{\kappa'}(R_q, K_q) = C(1 - \alpha_q)$. It is evident that these probabilities never depend on neither $\beta_q$ nor $\beta'_q$. Thus, again equality (4.6) is satisfied.

Having shown that (4.6) holds true for each pattern $R \subseteq Q$ and each state $K \in \mathcal{K}$ we conclude that $\kappa$ and $\kappa'$ are equivalent probabilistic representations of $(Q, \mathcal{K})$. $\square$

In the previous section a procedure is described for canceling out the $\alpha_q$ parameter of some item $q$ in an arbitrary knowledge structure. Concerning the $\beta_q$ parameters, the procedure follows similar lines. It is just matter of replacing the transformation in (4.8) with the following one:

$$\mathcal{K}' = \mathcal{K} \cup \{K \cup \{q\} : K \in \bar{\mathcal{K}}_q\}, \tag{4.10}$$

that does the correct job in transforming knowledge structure $\mathcal{K}$ into one which is forward-graded in $q$. Of course in this case Theorem 3 will be applied.

It is quite nice that the procedures described in this and previous sections can be applied to any knowledge structure, in principle. Nonetheless the question remains open whether the transformations (4.8) and (4.10), required by the two procedures, do make sense in any empirical context. Namely, if $\mathcal{K}$ is a knowledge structure that describes or respects some precise psychological construct or theory, it is likely that this theory will not be preserved anymore, after one of the two transformations (4.8) or (4.10) is applied. One would like to know, at least, which specific modifications the theory will undergo after their application. Precise answers to this question can be found if the knowledge structure $\mathcal{K}$ is delineated by some skill map.

### 4.3.4 Application to skill maps

In this section we show that, when a knowledge structure corresponds to some skill map, there is a precise way to make it backward-graded in some item $q$, thus forcing $\alpha_q$ to be zero. It is just matter of adding a new skill which is *specific* for that item. There is also a way to make it forward-graded in $q$: just add a new competency containing only a specific skill for $q$.

Consider a skill multi map $(Q, S, \mu)$, where $Q$ is the set of items, $S$ is the set of skills, and $\mu : Q \to 2^{(2^S)}$ is the skill assignment function. Let moreover $\mathcal{K}$

be the knowledge structure delineated by $\mu$, and assume that the states in $\mathcal{K}$ are delineated by $\mu$ through the competency model $F_\mu : 2^S \to 2^Q$ such that , for each $X \subseteq S$

$$F_\mu(X) := \{q \in Q : M \subseteq X \text{ for some } M \in \mu(q)\}.$$

Suppose to consider a new skill $s \notin S$ and to add it to $S$, obtaining thus a new collection $S' = S \cup \{s\}$ and a new skill map $(Q, S', \mu')$. Suppose, in particular that the function $\mu'$ is defined in the following way: having fixed a particular item $q \in Q$, for each item $i \in Q$

$$\mu'(i) = \begin{cases} \mu(i) & \text{if } i \neq q \\ \\ \{M \cup \{s\} : M \in \mu(i)\} & \text{if } i = q. \end{cases}$$

According to this definition, the skill map $\mu'$ differs from $\mu$ for the only fact that skill $s$ has been added to (each competency of) item $q$, while remaining identical to $\mu$ for any other item. This means that now $q$ cannot be solved by a student if she does not master skill $s$.

The knowledge states delineated by $\mu'$ are obtained by an application of the function

$$F_{\mu'}(X) := \{i \in Q : M \subseteq X \text{ for some } M \in \mu'(i)\}$$

to the subsets $X \subseteq S'$. Let $\mathcal{B} := \{F_{\mu'}(X) : X \subseteq S'\}$ be the knowledge structure delineated by $\mu'$. Then we have the following

**Observation 1.** The knowledge structure $\mathcal{B}$ is backward-graded in $q$.

*Proof.* For each $X \subseteq S$ the following two conditions hold true: (1) $F_{\mu'}(X) = F_\mu(X) \setminus \{q\}$, and (2) $F_{\mu'}(X \cup \{s\}) = F_\mu(X)$. Moreover, $\mathcal{K} = \{F_\mu(X) : X \subseteq S\} = \{F_{\mu'}(X \cup \{s\}) : X \subseteq S\}$, hence $\mathcal{K} \subseteq \mathcal{B}$. Furthermore let $q \in K \in \mathcal{B}_q$. Then it holds that $K = F_{\mu'}(X \cup \{s\})$ for some $X \subseteq S$ and hence $K = F_\mu(X)$, meaning that $K$ is in $\mathcal{K}$. That is $\mathcal{B}_q \subseteq \mathcal{K}$. Then we have the following chain of statements: $K \in \mathcal{B}_q$ implies $K \in \mathcal{K}$, which is true if and only if there exist $X \subseteq S$ such that $F_\mu(X) = K$, iff there exists $X \subseteq S$ such that $F_{\mu'}(X) = K \setminus \{q\}$, iff $K \setminus \{q\} \in \mathcal{B}$. $\square$

Observation 1 shows that adding a new skill $s$ to some item $q$ makes the knowledge structure backward-graded in $q$. Since $s$ is only added to item $q$, the interpretation is that $s$ is a *specific skill* for $q$. In a practical application, Theorem 1 and Observation 1 can be jointly applied in order to (i) identify items whose $\alpha$ probability is too high to be credible as a careless error and (ii) rectify the model by adding a specific skill to each of those items.

Suppose now that another skill map $(Q, S', \mu'')$ is constructed, which respects the following rule for every item $i \in Q$:

$$\mu''(i) = \begin{cases} \mu(i) & \text{if } i \neq q \\ \mu(i) \cup \{\{s\}\} & \text{if } i = q \end{cases}$$

In this skill map a new competency $\{s\}$ is added to item $q$, while the rest is left unchanged. Let $\mathcal{F}$ be the knowledge structure delineated by $\mu''$. Then we have

**Observation 2.** The knowledge structure $\mathcal{F}$ is forward-graded in $q$.

*Proof.* For $X \subseteq S$ the following two conditions hold true: (1) $F_{\mu''}(X) = F_{\mu}(X)$, and (2) $F_{\mu''}(X \cup \{s\}) = F_{\mu}(X) \cup \{q\}$. Moreover $\mathcal{K} = \{F_{\mu}(X) : X \subseteq S\} = \{F_{\mu''}(X) : X \subseteq S\}$, therefore $\mathcal{K} \subseteq \mathcal{F}$. Furthermore: $K \in \bar{\mathcal{F}}_q$ implies $K \in \mathcal{K}$, which is true iff $F_{\mu}(X) = K$ for some $X \subseteq S$ iff $F_{\mu''}(X) = K \cup \{q\}$ for some $X \subseteq S$ iff $K \cup \{q\} \in \mathcal{F}$. $\square$

Concerning the applicability of the procedure described above, there might be specific issues that arise in connection to the type of probabilistic model that one decides to apply (the BLIM, rather than some other special case of it). For instance, when the unrestricted BLIM is applied, one has to consider that there is one parameter $\pi(K)$ for each knowledge state in the model. Making the knowledge structure backward-graded (or forward-graded) with respect to some item $q$ will, most likely, add a quite large number of new states (and thus parameters) to the model. Apart from considerations about the available degrees of freedom, model complexity is an issue that cannot be overlooked in this case.

For the type of probabilistic models introduced here, under the condition of skill independence adding a new skill to an item $q$ means adding exactly one new parameter (the skill probability) to the model. However, since this is compensated by constraining $\alpha_q$ to zero, the number of degrees of freedom remains unchanged. Similar conclusions can be drawn for the introduction of a new competency for an item. Some different considerations will be carried out in this respect when referring to models with more than one latent class: in that case the introduction of a new skill in the model means subtracting a number of degrees of freedom equal to the number of latent classes minus 1. Finally the question remains open of how large a careless error or a lucky guess should be for making the introduction of a new skill or new competency worthwhile. In general, provided that the condition $\alpha_q < 1 - \beta_q$ is respected, a value of $\alpha_q$ (respectively, $\beta_q$) greater or equal to .5 should provide strong evidence in favor of the existence of a specific skill (resp., competency) for item $q$ but, depending on the empirical context, smaller upper bounds could also be considered.

## 4.4 An application example

We are now going to present an application example of the method introduced above. More specifically we will refer to the construction of a knowledge structure, starting from a skill map which represents the presence of clinical symptoms (attributes) into the set of items composing the sub-scale Cleaning of the MOCQ-R. The sub-scale includes 8 items investigating the specification of the OCD related to the fear of contamination and the compulsive activity of cleaning. The approach and the application context are similar to those in Spoto, Stefanutti and Vidotto (2010) and Spoto et al. (2008).

In our application we started from the construction of a conjunctive skill map associating to each item $i$ a set of diagnostic criteria (attributes) for the OCD. We then built a matrix where each row represents an item and each column represents a skill. We analytically assigned to each item a subset of skills (i.e. the clinical characteristics investigated by that item).

The special case of the BLIM presented in section 4.2 has been used as the reference model. An iterative MatLab procedure applies the Expectation Maximization algorithm (Dempster et al., 1977) for the estimation of the model parameters by maximum likelihood. Furthermore the goodness of fit statistics Chi-square has been computed and here presented along with the condition number (an index of identifiability of the model). For an outline of the algorithm involved in the test of the model see Appendix (7.2).

This procedure allows several interesting considerations both from a methodological and clinical perspective. In fact, on the one hand the general fit indexes could provide us with an overview of the quantitative properties of the hypothesized model; on the other hand both the error parameters (i.e. $\alpha$ and $\beta$) and the information about each single skill represent a unique tool to evaluate the validity and reliability of the items and of the structure. Moreover the estimated probability of every single skill may be informative about the opportunity of keeping that skill in the model. Given that the estimate of every single skill lies in the open interval $(0, 1)$, a specific skill with an estimated probability of, say .99 (every students present this skill) has no practical use; therefore the skill should be removed from the model. On the other hand a small skill probability would be indicative of the presence of a quite difficult or rare skill.

The models have been tested using a large database of 4412 individuals who filled the MOCQ-R as a part of a broad spectrum assessment performed using the CBA 2.0. Individuals were adult males and females from Northern Italy. The sample is supposed to represent the normal population where the prevalence datum for OCD is approximately the 3%.

In the present section we will describe several different steps of the performed procedure. At each step a modification in the skill map and/or in the parameter estimates is introduced.

### 4.4.1 The Conjunctive Skill Map

The starting point of the application was the conjunctive skill map displayed in Table 4.1, where each item was associated with a single set of skills. In that Boolean matrix each row represents an item and each column represents a skill. A 1 in a cell $(i, s)$ indicates that the attribute $s$ is investigated by item $i$. In table 4.2 all the parameters estimates are displayed . From this

**Table 4.1:**

The conjunctive skill map.

| Item/Attribute | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $a_9$ |
|---|---|---|---|---|---|---|---|---|---|
| $i_1$ | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_2$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| $i_3$ | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| $i_4$ | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| $i_5$ | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| $i_6$ | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| $i_7$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| $i_8$ | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |

table it is possible to make some considerations. First, it can be seen that the model presents a very high value of Chi-square. It is worth explaining

**Table 4.2:**

Fit indexes and error parameters estimates for the first conjunctive model with $N = 4412$, $\chi^2(230) = 1300.35$; $p < .001$ and condition number= 6.422.

| Item | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | $i_6$ | $i_7$ | $i_8$ |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.809 | 0.474 | 0.198 | 0.414 | 0.102 | 0.001 | 0.645 | 0.289 |
| $\beta$ | 0.000 | 0.000 | 0.654 | 0.442 | 0.001 | 0.070 | 0.000 | 0.328 |

this point because it could be considered as an indication for the misfit of the model. However, it is a well known fact that this statistic is very sensible to the sample size. In the present application we have a very high number of subjects, thus the Chi-square statistic value could be actually affected by sample size. Further information arises from the error parameters estimates. It is evident how there are several problems with many items. This is a strong indication about the need to reformulate the model by accounting for these elements.

In order to do so, we first note that in the matrix presented in Table 4.1 there are two items requiring specific attributes. Thus, following the procedure illustrated in the previous sections we can deflate the $\alpha$ parameters for those items. Besides, this method should also solve identifiability problems.

### 4.4.2 Managing Careless Error Estimate I

As shown in Table 4.1, items 3 and 6 have specific attributes. The second step of this application example consisted of fixing to 0 (and fixing them in the estimate algorithm) the values of the $\alpha$ parameters for these two items. The fit indexes and the parameters estimate for this second model are displayed in table 4.3. It can be easily seen that no major improvements arise from

**Table 4.3:**

Fit indexes and error parameters estimates for the second conjunctive model obtained by managing the careless error of item 3. The model was tested using a sample of 4412 subjects, $\chi^2(230) = 1300.69$; $p < .001$ and condition number= 6.609.

| Item | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | $i_6$ | $i_7$ | $i_8$ |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.809 | 0.474 | 0.000 | 0.414 | 0.096 | 0.000 | 0.645 | 0.289 |
| $\beta$ | 0.000 | 0.000 | 0.654 | 0.442 | 0.001 | 0.070 | 0.000 | 0.328 |

this first modification. This happens because even in the previous model the values of the inflated parameters were actually very low. In other words the problems with the model (from the careless error perspective) are mostly due to the high values of the error rates of items 1, 2, 4 and 7. Thus the skill map has to be modified in order to cope with the evidence of the fact that there are some diagnostic characteristics investigated by these items that are not included in the model.

### 4.4.3 Managing Careless Error Estimate II

A reformulation of the skill map was conducted in order to cope with the above described issues. The new matrix is displayed in table 4.4. Specific

**Table 4.4:**

The conjunctive skill map obtained by adding specific skills (attributes) to items 1, 2, 4, 6 and 7.

| Item/Attribute | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $a_9$ | $a_{10}$ | $a_{11}$ | $a_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i_1$ | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_2$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| $i_3$ | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_4$ | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| $i_5$ | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_6$ | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| $i_7$ | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| $i_8$ | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

attributes have been introduced for the listed items. Furthermore, it has to be stressed that the present structure is the final step of a list of successive modifications. Some refinements in the specification of each item' set of skills have been performed. The results of the parameter estimates for this model are displayed in table 4.5. From the table we can extract information about

**Table 4.5:**

Fit indexes and error parameters estimates for the third conjunctive model obtained by managing many careless errors and adding specific attributes for some items. The model was tested using a sample of 4412 subjects, $\chi^2(227) = 1240.20$; $p < .001$ and condition number= 3.215.

| Item | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | $i_6$ | $i_7$ | $i_8$ |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.000 | 0.000 | 0.274 | 0.000 | 0.293 | 0.000 | 0.000 | 0.274 |
| $\beta$ | 0.028 | 0.000 | 0.656 | 0.432 | 0.000 | 0.080 | 0.007 | 0.354 |

the main issues introduced in the previous sections. If we compare the Chi-square value observed for this model with the one displayed in tables 4.2 and 4.3 we immediately notice that it is lower. It has to be stressed how the difference is not dramatic, but, considering the high number of subjects and the condition number (that is actually a half of the previously observed ones), it can be considered as reliably lower. Some further evidence supporting the better goodness of this model arise from the error rates evaluation. In fact it can be seen how no further problems exist for the $\alpha$ estimates, while some issues are still present with respect to the lucky guess estimates. It is worth to remind here that in our specific field of application it could be a nonsense to refer to $\alpha$ and $\beta$ as respectively *Careless Error* and *Lucky Guess*. A more adequate interpretations of these parameters could be respectively *False Negative* and *False Positive* (Spoto, Stefanutti, & Vidotto, 2010; Spoto et al., 2008). Given this particular meaning of the parameters, the response format of the questionnaire (i.e. dichotomous), and the fact that the items under analysis are clinical (i.e. differently form the traditional fields of application of KST, some further aspects about the responding process have to be taken into account such as simulation, low level of insight, social desirability, etc.), it seems reasonable to accept a higher value of error rates that, anyway, should not be higher than 0.40.

### 4.4.4 Managing Lucky Guess Estimates: the Skill Multi Map

The last step of the present example consists of the construction of a skill multi map able to cope with the high values of the $\beta$ estimates. By introducing more than one subset of attributes alternatively needed to answer an item, it has been proven that a model exists whose $\beta$ parameter for that item is equal to 0. The resulting multi map is displayed in table 4.6. It has to be stressed that this solution is not surprising in a clinical context, where

**Table 4.6:**

The skill multi map.

| Item/Attribute | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $a_9$ | $a_{10}$ | $a_{11}$ | $a_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i_1$ | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| $i_2$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| $i_3$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_3$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_3$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_3$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_4$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_4$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_4$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_4$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_5$ | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_6$ | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| $i_7$ | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| $i_8$ | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $i_8$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $i_8$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

several sets of symptoms could lead to the same diagnostic category. This is because of both the complexity of the psychological assessment itself and of the different symptoms actually noticed by the patients that can vary for the same disorder from a patient to another. Thus the answer to an item could be obtained for different reasons that a multi map seems more adequate to capture and describe. The results obtained for this last model are displayed in table 4.7. Notice that the overall fit indexes are better than those ob-

**Table 4.7:**

Fit indexes and error parameters estimates for the third conjunctive model obtained by managing many careless errors and adding specific attributes for some items. The model was tested using a sample of 4412 subjects, $\chi^2(227) = 1240.20$; $p < .001$ and condition number= 3.215.

| Item | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | $i_6$ | $i_7$ | $i_8$ |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.000 | 0.000 | 0.284 | 0.000 | 0.293 | 0.000 | 0.000 | 0.374 |
| $\beta$ | 0.040 | 0.000 | 0.000 | 0.000 | 0.007 | 0.039 | 0.006 | 0.274 |

served for the previous model. Nonetheless, it has to be stressed that the Chi-square value is still high. One possible explanation of this is that all the individuals are included in the same class. It seems reasonable to assume that in the population at least two latent classes of subjects exist: the normal and the clinical. Thus a potential improvement of the model could be the introduction of a latent classes structure in order to better discriminate response patterns. Finally it is observed how none of the error rates is higher than 0.40.

A final remark can be referred to the single skills' probabilities displayed in table 4.8. It can be observed how they range between 0.35 to 0.93. This

**Table 4.8:**

Attributes' probabilities for the estimated model.

| Attr | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $a_9$ | $a_{10}$ | $a_{11}$ | $a_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0.35 | 0.39 | 0.60 | 0.63 | 0.64 | 0.45 | 0.86 | 0.46 | 0.50 | 0.63 | 0.93 | 0.51 |

means that there are some clinical characteristics that can be easily found in the general population, but that standing alone they may not be an index of a psychological problem. On the contrary, the detection in a subject of less likely characteristics could be an index of some clinical issues.

## 4.5 Introduction of more than one latent class

The models presented in the previous section consider only one latent class. This could be an oversimplification of the situation to be described. More specifically, when referring to a set of clinical items it can be easily argued that the answering behavior could be, at least, roughly split into two latent classes: the clinically significant on the one hand and the not-clinically significant on the other hand. Moreover, it has to be stressed how these two classes are not supposed to lye on the same continuum, but they are located on a multidimensional space. This particular aspect is well described by the probability estimates of each single attribute in each class. If we were on a continuum the clinical attributes investigated by the items of the questionnaire are supposed to present a higher probability to occur into the clinical class than in the non-clinical one. On the contrary we will show in a moment how we observed some attributes with a higher probability in the non-clinical class.

In order to estimate the model with more than one latent class we used the same algorithm run for the previous analysis. The skill map to be tested was the one displayed in Table 4.6, and the same $\alpha$ and $\beta$ parameters above have been fixed to 0.

The error rates estimate are displayed in Table 4.9. What immediately emerges is the dramatic reduction of the Chi-square value. This suggests that the phenomenon to be described is actually better depicted by a two latent classes solution. On the other hand, the increment of the condition number suggests a worse identifiability of the model. Nevertheless, this index seems to fall into an acceptable boundary, thus the model can be reliably accepted.

**Table 4.9:**

Fit indexes and error parameters estimates for the third conjunctive model obtained by managing many careless errors and adding specific attributes for some items. The model was tested using a sample of 4412 subjects with 2 latent classes, $\chi^2(214) = 388.77$; $p < .01$ and condition number= 127.591.

| Item | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | $i_6$ | $i_7$ | $i_8$ |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.000 | 0.000 | 0.246 | 0.000 | 0.007 | 0.000 | 0.000 | 0.339 |
| $\beta$ | 0.052 | 0.020 | 0.000 | 0.000 | 0.209 | 0.082 | 0.003 | 0.280 |

As previously introduced, using this kind of model it is possible to estimate the probability of each single attribute in each latent class and the probability of the class itself. For simplicity here we will call the two classes "clinical" and "not-clinical", but it can be argued that it would be possible to have as many classes as the number of disorders to be investigated. In this way we will be able to estimate the probability to observe a response pattern of each class and the probability of each attribute in each class.

In the tested model we have that the probability of the "clinical" class is 0.627, while the probability of the "not-clinical" class is 0.373. Probabilities of each single attribute in each class are reported in Table 4.10. It is very important to note the particular dimensionality represented by the two latent classes of the model. If we were on a continuum the expected trend

**Table 4.10:**

Attributes' probabilities for the estimated model (C= Clinical; NC= Not-Clinical) .

| Attr | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $a_9$ | $a_{10}$ | $a_{11}$ | $a_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $NC$ | 0.02 | 0.06 | 0.35 | 0.03 | 0.65 | 0.22 | 0.41 | 0.12 | 0.80 | 0.55 | 0.19 | 0.22 |
| $C$ | 0.60 | 0.98 | 0.98 | 0.59 | 0.99 | 0.64 | 0.69 | 0.79 | 0.54 | 0.25 | 0.66 | 0.48 |

of attributes' probabilities would be increasing from the "non-clinical" to the "clinical" class. What we actually observe is that not all the attributes present that specific trend. More specifically, it can be seen how attributes 9 and 10 have a higher probability value in the "non-clinical" class. This means that the two classes are located in a multidimensional space. The number of dimensions could range from context to context, but hypothetically it could include as many classes as the number of attributes. Therefore, the introduction of more latent classes could allow the identification of specific diagnostic configurations depicting a variety of disorders.

## 4.6 Discussion

The present chapter was mainly aimed at presenting a method to deflate $\alpha$ and $\beta$ probabilities for some items in a PKS obtained out of a skill multi map. It has been proved how a high value of $\alpha$ indicates the presence of one specific attribute referred to that item, while a high value of $\beta$ suggests the presence of multiple solution strategies for that item. Furthermore, it has been shown how such parameters could be used as diagnostic tools to improve the skill map model.

A probabilistic model for PKS, accounting for multiple solution strategies of different items, has been formalized and applied to the KS obtained through a skill multi map. Finally an application example of the potential empirical utility of the proposed model has been presented on a set of clinical items.

Further developments of the proposed methodology could allow the construction of an adaptive algorithm for the assessment of the so called *clinical state* (Doignon & Falmagne, 1999) of individuals. Moreover, from the formal point of view, it will be investigated whether the extension of the present model to a more general class of KS, i.e. not necessarily graded in an item, is feasible.

Another interesting perspective of the present approach is represented by the potential definition of a set of latent classes to describe the multidimensional space of the psychological disorders. It has been shown how this can be performed with two rough classes, but further developments could include three or more classes in order to depict the more in more accurately the space under investigation.

# Chapter 5

# Knowledge Structure Construction Through Database Query

## 5.1 Entail Relation and Query Procedure

The subject matter of this chapter is the description of a third way to build a knowledge structure to be used in the clinical assessment. In the previous chapters we referred to a set of techniques involving the cognitive concepts underlying the behavioral process of answering an item. These cognitive concepts were in one case the attributes of a formal concept (chapter 3), and, in the other case, the skills of a skill multi map (chapter 4). In the present chapter we are going to introduce a new procedure based on a more "traditional" approach in KST. In fact, the most popular way to build a knowledge structure in literature is the so called "Query to expert" procedure (Falmagne et al., 1990; Koppen & Doignon, 1990; Düntsch & Gediga, 1996; Koppen, 1998; Doignon & Falmagne, 1999). Using this procedure an *entail relation* (Doignon & Falmagne, 1999; Burigana, 2004; Falmagne & Doignon, 2010) is defined and used in order to identify the potential admissible knowledge states of the structure out of the power set of the domain.

**Definition 4.** Given a knowledge space $\mathcal{K}$ defined on the domain $Q$, a relation $\mathcal{P}$ defined between $2^Q$ and $Q$ such that for any $X \subseteq Q$ and any item $q \in Q$

$$X\mathcal{P}q \text{ iff } (\forall K \in \mathcal{K} | X \cap K = \emptyset) \Rightarrow q \notin K \tag{5.1}$$

is called entail relation.

The interpretation of the entail relation as defined above is: "failing all the items in $X$ entails failing $q$". The test of this kind of implications through the administration to experts of dichotomous questions is the "query to experts procedure". Questions are formulated like: "Suppose that an individual has just failed all the items included in the set $X$; is it almost certain that he will fail even item $q$? Suppose that the assessment is carried out in optimal conditions without either guessing or careless". The interrogation starts with $|X| = 1$ and stops when no questions can better define the entailment among the items of $Q$.

A one to one correspondence is defined between the family $\mathcal{K}$ of knowledge spaces on $Q$ and the family $\mathcal{P}$ of entail relations for $Q$ (Koppen & Doignon, 1990). This correspondence is defined through equation 5.1 and the following:

$$K \in \mathcal{K} \text{ iff } (\forall (Q, q) \in \mathcal{P} | X \cap K = \emptyset) \Rightarrow q \notin K. \tag{5.2}$$

It follows that any knowledge space on $Q$ can be recovered from its own entailment $P$.

Several methodologies and algorithms have been studied and created throughout the years to address the completion of a query and, in the end, the construction of a knowledge structure. Some of these procedures start from the definition of the entailment on the domain and out of this derive the knowledge structure (Koppen & Doignon, 1990; Koppen, 1993; Stefanutti & Koppen, 2003); some other suggestions start from the definition of different kinds of relations on the knowledge domain (Dowling, 1993; Albert & Held, 1994; Düntsch & Gediga, 1995; Cosyn, 2002).

One of the main issues concerning the query to expert procedure is related to the extreme consumption of time. In order to cope with this issue, several inference rules have been derived in order to avoid redundant queries.

Furthermore, it has been investigated a way to address the problem of integrating different experts' structures (Dowling, 1994; Schrepp & Held, 1995; Schrepp, 2001).

The advantages of this kind of procedures are mostly related to the fact that the relations among items are defined, in theory, independently from the skills needed to master each item and to the fact that the question posed to the expert is well accepted in this kind of context. The limitations are due to inferential fallibility of the expert that can be only partially corrected through the inference rules and through the application of the non-contradiction criterium. Furthermore this kind of queries build a-priori structures to be tested on real subjects.

Here we will introduce a methodology representing an improvement of the construction of a knowledge structure starting from answers collected on a set of subjects. This kind of procedure have been already carried out with different algorithms (Doignon & Falmagne, 1999; Schrepp, 1999; Cosyn & Thiéry, 2000), but the algorithm we used is pretty new and under experimentation at the ALEKS corporation. An outline of this algorithm is presented in the form of a draft in Falmagne and Doignon (2010).

## 5.2 Database Query Algorithm and Learning Spaces

The most relevant existing query procedures are based on "expert query". In this section we introduce a different approach to query based on the interrogation of databases of data previously collected on a set of items. We will refer to a set of clinical items (namely the items of the MOCQ-R used in the previous chapters) administered to the sample of persons introduced in Chapter 4. This large data set will allow both the investigation of the relations among the clinical items at issue and the recovery of a knowledge structure from the derived entailment.

In order to build the entailment a specific algorithm has been created. The logic underlying this procedure is very similar to the one used by "traditional" queries. In other words, the interrogation is carried out through the same kind of questions like:

*Suppose that a student under examination has just failed all the items in a set $X$ (or is regarded as not capable of solving any of these items). Is it practically certain that this student will also fail item $q$? Assume that the conditions are ideal in the sense that errors and lucky guesses are excluded.*

This question, in our application, is not posed to an expert, but it is checked within the observed response patterns. The procedure starts by comparing pairs of items and proceeds by increasing the cardinality of $X$. In order to accept or discard the relation $A\mathcal{P}q$ a decision rule has to be chosen. This rule has to take into account the fact that a part of noise is included in any sample of data. On the other hand the rule has to be able to minimize both false positive and false negative entailments. The entailments in the data set not satisfying the defined criteria are excluded and the structure is built out of the remaining queries.

Four criteria have been chosen in order to identify the reliable queries: forward conditional probability ($\beta_1$), backward conditional probability ($\beta_2$), composed conditional probability ($\beta 3$) and the error rate ($\beta 4$). These indexes were calculated as follows[1]. In the formulas below for each string of three letters the first two are referred to the answer to the items included in $X$, while the third one is referred to the answer to $q$. For instance the string $TTF$ indicates that both the items in $X$ present a "True" answer, while $q$ is "False". These formulas are computed on a data set composed by the so called open queries including a number of columns equal to the possible

[1]Notice that for simplicity here are presented the formulas regarding the case in which $|X| = 2$. This means that the entailment is defined between the first two items and the third one.

combinations of answers (in the presented example these columns are 8, i.e. $TTT$, $TTF$, $\ldots$, $FFF$), and as many rows as the number of open queries.

$$\beta_1 = \frac{TTT + TFT + FTT}{TTT + TFT + FTT + FFT} \tag{5.3}$$

$$\beta_2 = \frac{FFF}{FFF + FFT} \tag{5.4}$$

$$\beta_3 = \frac{TTT + TFT + FTT + FFF}{TTT + TFT + FTT + 2FFT + FFF} \tag{5.5}$$

$$\beta_4 = \frac{FFT}{TTT + TTF + TFT + TFF + FTT + FTF + FFT + FFF} \tag{5.6}$$

A very short introduction of the four indexes is needed:

- Forward Conditional Probability ($\beta_1$): it is the ratio between the number of times the entailment between the items in $X$ and $q$ is satisfied when a positive answer to $q$ is observed, and the number of times that $q$ presents an affirmative answer. In fact, the condition $FFT$ actually falsifies the entailment since a positive answer is observed to $q$ even if none of the items in $X$ has an affirmative answer;

- Backward Conditional Probability ($\beta_2$): it is the ratio between the number of times the entailment between the items in $X$ and $q$ is satisfied when a negative to $q$ is observed, and the number of times all the items of $X$ present a negative answer. Once again the combination $FFT$ represents a violation of the entailment;

- Composed Conditional Probability ($\beta_3$): it simply puts together the first two formulas in order to obtain an index addressing for both the conditions above;

- Error Rate ($\beta_4$): it represents the proportion of times the entailment is violated.

The values of the three conditional probabilities (i.e. $\beta_1$,$\beta_2$ and $\beta_3$) are expected to be as high as possible while the error rate (i.e. $\beta_4$) is expected to be close to zero. Thus the selection rule for the queries was fixed as follows: $\beta_1$, $\beta_2$ and $\beta_3 > 0.90$; $\beta_4 < 0.05$. At each block of the procedure these four indexes were calculated and on the basis of their value the corresponding query was either included or discarded in the definition of the structure. When no query was left open, the algorithm stopped and the structure was built.
In the specific case presented in this chapter, the algorithm was modified in order to allow the construction of a *learning space* (LS; Falmagne & Doignon, 2010; Cosyn & Uzun, 2009; Falmagne, Cosyn, Doignon, & Thiéry, 2006). Some theoretical and descriptive concepts about LS are introduced below. From a merely formal standpoint a LS is a knowledge structure ($\mathcal{K}$) satisfying the following two axioms:

1. the family $\mathcal{K}$ is *well-graded*, that is, if $K$ and $L$ are two states by $n$ items, then it exists a chain of states $K_0 = K, K_1, \ldots, \mathcal{K}_n = L$ such

that for $0 \leq i \leq n$, the two states $K_{i+1}$ and $K_i$ differ by exactly one item;

2. The family $\mathcal{K}$ is closed under union, i.e. $\mathcal{K}$ is a knowledge space.

These axioms are very important form a theoretical point of view. Assuming e LS model implies accepting that there always exist a path connecting two states by adding one item at a time. In other words it means that given two response patterns (two patients) they can be connected through a sequence of steps including one item each. It is evident how this assumption is strong and how it drastically changes the perspective compared to what introduced in the previous chapters. Here we aim at introducing a method and we leave to further investigations the role and the specific fields of application of LS in the clinical context.

We can now summarize the steps of the procedure:

1. A database including the answers to the MOCQ-R (20 items) of a sample of 4412 individuals has been used;

2. The database query has been performed using the introduced algorithms;

3. After each step of the query (with increasing cardinality of $X$) the indexes $\beta_1$,$\beta_2$, $\beta_3$ and $\beta_4$ have been calculated;

4. The cut-offs were set at: 0.90 for $\beta_1$,$\beta_2$ and $\beta_3$; at 0.05 for $\beta_4$;

5. Entailments not satisfying all these parameters were discarded while those falling into the acceptance range were included in the LS definition;

6. When no query was left open, the algorithm stopped and the LS was generated.

## 5.3 Results and Discussion

The result of the database query procedure is essentially a LS representing the implications among the items of the questionnaire given the reliable entailments found. Some interesting observations can be carried out on the output of the procedure.

First of all it is important to note how this procedure actually succeeded in generating a structure. This result is not that trivial. In fact, it has to be recalled that, in this case, no theoretical structure was hypothesized, but the analysis was only on a merely behavioral level. In other words, even if many clinician could not identify any prerequisite relation among items (it is extremely difficult to say how a clinical item could entail one or many different items), this relation actually is present among the items, and this relation is detected and tested by the construction algorithm. Thus it is possible to identify a structure underlying the MOCQ-R. In general, if the situation were of no structure among items, the number of the admissible states would equal the cardinality of the powerset of the set of items. In the application at hand this number would be $2^{20}$. The important result obtained from the application of the database query algorithm is that only 10991 states are admissible for the investigated domain. This number is approximately the 1% of the powerset's cardinality ($2^{20} = 1048576$). It has to be noticed how the entailments included in the LS definition either satisfy all the criteria $\beta_1$,$\beta_2$, $\beta_3$ and $\beta_4$, or are direct inferences following the no-contradiction principle and the satisfaction of the LS axioms.

A second important remark about the results has to be referred to the fact that, actually, the axioms of LS are satisfied by the used database. This means that: both the empty set and the total set are states of the structure; the structure is closed under union (these two characteristics define that what has been derived is a knowledge space); the structure is well graded. This last characteristic introduces a number of critical aspects to be in depth investigated, hopefully by further studies. For example it has to be discussed whether it is meaningful to make all these assumptions; it has to be checked in detail what entailments are the reliable ones, and how they can be explained from a theoretical point of view. Furthermore it has to be deeply investigated the probability of each state of the LS. Nevertheless, the proposed application demonstrates the formal applicability of these models to the behavioral process of response to a set of clinical items.

Finally it has to be stressed how, given the characteristics of the approach adopted in this application, we do not have any information about the false positive and false negative parameters, nor about the probability of eventual latent classes or attributes. This aspect has to be taken into serious account in a clinical context where the psychologist is mostly interested in understanding what characteristics a patient demonstrates to present, rather than the score or the response pattern itself. On the other hand, it is possible here to manage a larger set of items without referring to specific sub scales, but delineating exactly the structure on the whole questionnaire. In fact, this kind of query is suitable for very large database collected on a wide set of items, the situation in which this procedure has been developed (i.e. at the ALEKS Corporation). Thus, this procedure is very useful when coping with databases referred to hundreds of items and many thousands of assessments. In that contexts items are built in order to fulfil a specific learning program (e.g. Beginning Algebra, Middle School Math Course, etc.), and no specific a priori skill assignment is useful. In fact, in that case, what a student knows is depicted in a pretty precise way by the items he is able to solve. Some different considerations could be carried out in the clinical context where an affirmative answer to an item does depend not only on the ability of the patient to individuate that specific characteristic, but also on its level of insight, its decision making threshold and so on. Furthermore, it can be said that each item in the ALEKS (learning) environment makes sense by itself (the information about what the student is able to solve is sufficient to plan the further steps of the learning process); on the other hand the information that a clinician is interested in collecting is not only to what items the patient answered positively, but also what diagnostic and therapeutic information that item provides by itself and together with different sets of items. In other words what is interesting for a psychologist is the attribute (the skill) behind the item. Nevertheless, the proposed procedure represents a very interesting starting point to solve the problem of representation of clinical structures and it can be hypothesized that the construction of an ex-post explanation of the entailments found on the items could represent a way to clarify most of the exposed criticisms.

# Chapter 6

# General Discussion

In the present dissertation we introduced the theoretical and mathematical foundations of FPA. The main aim of this psychological assessment methodology is to apply a set of formal mathematical rules to the clinical assessment context. What traditional test theories such as CTT, and IRT take into account are measurement-psychometric characteristics mostly related to the score an its reliability. None of these two approaches account either for the specific characteristics investigated by each single item, or for the specificity of each single response pattern. FPA is aimed at introducing an approach able to address all these issues through an adaptive algorithm. Along the thesis we shown how the approach we are proposing could actually provide a clinician with: a higher amount of information out of a set of questions; a higher measurement reliability and validity; the opportunity to process in a faster way a higher number of information for both horizontal and vertical integration.

It has been described how the two mathematical reference point of FPA are KST and FCA, while its clinical theoretical forerunner is the CBA 2.0 battery. From a clinical perspective the most relevant innovation introduced by FPA is the possibility to describe each clinical item as a collection of investigated diagnostic symptoms. This aspect is even more interesting if referred to the fact that, in general, it is possible to describe the same item through different attribute configurations based on the theoretical background we want to adopt. In other words, even if we defined the set of diagnostic criteria of the DSM IV-TR investigated by each item of the MOCQ-R, someone else could be interested in a more cognitive-behavioral solution of the problem by identifying what theoretical aspects are explored by each single item. Another important innovation introduced by FPA in the clinical framework is the possibility to identify formal relations among both diagnostic criteria and clinical items. This opportunity was trivial for the traditional fields of application of KST and FCA, but it is absolutely an important finding for the clinical context. Some interesting findings based both on the skill maps and on the significance of the response patterns have been displayed. For instance it can be seen in Figure 3.3 how the three response patterns on the top of the lattice (namely nodes 2, 4 and 3) are at the same level of complexity in the structure, but one of them (node 3) corresponds to a score of 5 that is not clinically significant for the norms of the questionnaire. This node represents a very clear example of the innovation introduced by FPA. In fact, it is evident that the number and quality of attributes that a patient is supposed to present when responding to the five items included in node 3 are as clinically relevant as those presented by a patient who scored 7 with a different set of attributes and (naturally) with a different response pattern. The third relevant innovation from a clinical perspective is the opportunity, introduced in chapter (4), to hypothesize and test the probability of each diagnostic attribute within a number of theoretical latent classes. Through this methodology it will be possible to estimate the probability of each clinical attribute in a set of different classes corresponding , for example, to a set of different diagnosis. This element is a really important improvement. In fact, what often happens now is that a clinician performs a test, finds it to be significant and then checks for the presence of the diagnostic criteria for a specific diagnosis. With FPA the clinician will submit the adaptive procedure and he will obtain the probability of the patient to be in a specific latent class, his probability to present each attribute and his probability to present a specific response pattern. It is trivial to understand the difference in quality and quantity of information.

From a formal mathematical point of view the most relevant innovations introduced by FPA consists of both a new way to connect FCA and KST (see chapter 3), and a methodology to manage the error parameters estimated in the BLIM. In particular, this second element represents an opportunity to better understand, even from a theoretical point of view, some properties of this kind of knowledge structures. In other words, a methodological solution has been found to cope with a theoretical problem related to the specificity of some elements included in some items. This methodological solution can be applied also when coping with different configurations of attributes underlying an affirmative answer to an item. The present thesis allowed also the collection of a wide body of information useful for the calibration of an algorithm able to carry out the clinical adaptive assessment. Among this information we recall here the $\alpha$ (false negative) and $\beta$ (false positive) error rates estimate, the probability of each latent class, the probability of each attribute in each class. Another important methodological remark has to be addressed to the "query to database procedure". This technique needs to be further investigate since it represents the fastest way to derive a knowledge structure. On the other hand it does not present the clinical properties revealed by the procedures proposed both in chapters (3) and (4). It would be interesting to mesh these three approaches in order to take the best out of each one and to compensate the limits they present individually.

The meshing of the three approaches represents one of the potential developments of FPA. Other interesting perspectives are, for instance, the actual construction of the adaptive algorithm, the study of the best form of output to be provided to the clinician who uses FPA, the application of the methodology to different pathological areas and the generalization of the findings about the management of the error parameters.

During these years of research we explored many different aspects of a very complex field, but it is evident how many other issues have still to be considered, studied and, eventually, implemented.

# References

Albert, D., & Held, T. (1994). Establishing knowledge spaces by systematical problem construction. In D. Albert (Ed.), *Knowledge Structures* (pp. 81–116). Berlin: Springer-Verlag.

Albert, D., Schrepp, M., & Held, T. (1992). Construction of knowledge spaces for problem solving in chess. In G. H. Fisher & D. Laming (Eds.), *Contributions to Mathematical Psychology, Psychometrics, and Methodology* (pp. 123–135). New York: Springer-Verlag.

American Psychiatric Association [APA]. (1995). *DSM-IV-TR. Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition, Text Revision ed.). Washington D. C.: American Psychiatric Association.

Arrindell, W. A., & VanDerEnde, J. (1985). Cross-sample invariance of the structure of self-reported distress and difficulty in assertivenes: experiences with the Scale for Interpersonal Behaviour. *Advances in Behaviour Research and Therapy*, *27*, 205–243.

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic. (2nd edition).* New York: Guilford Press.

Barlow, D. H., & Durand, V. M. (2004). *Abnormal Psychology: An Integrative Approach.* Belmont, CA: Thomson Wadsworth.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for Beck Depression Inventory II (BDI-II).* San Antonio, TX: Psychology Corporation.

Bellak, L., & Bellak, S. S. (1949). *Children's Apperception Test.* Oxford, England: CPS Co.

Bertolotti, G., Zotti, A. M., Michielin, P., Vidotto, G., & Sanavio, E. (1990). A computerized approach to cognitive behavioural assessment: An introduction to CBA-2.0 primary scales. *Journal of Behavior Therapy and Experimental Psychiatry*, *21 (1)*, 21–27.

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

Birkhoff, G. (1967). *Lattice Theory.* Providence, R.I.: American Mathematical Society Colloquium Pubblication. (No. XXV)

Boole, G. (1976). *Indagine delle leggi del pensiero, su cui sono fondate le teorie matematiche della logica e della probabilit.* Torino: Einaudi. (Reprinted from *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*, 1854)

Boole, G. (1993). *L'analisi matematica della logica.* Torino: Bollati Boringhieri. (Reprinted from *The mathematical analysis of logic, being an* *essay towards a calculus of deductive thinking*, by Barclay & MacMillan, Eds., 1847, Cambridge, UK)

Burigana, L. (2004). *Vedute di psicologia matematica in prospettiva discreta.* Padova: CLEUP.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Perry, J. N., & Atlis, M. M. (2000). Validity and utility of computer-based test interpretation. *Psychological Assessment*, *12 (1)*, 6–18.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, *31*, 141–154.

Corman, L. (1969). *Le Test PN, Manuel.* Presses Universitaire de France.

Cosyn, E. (2002). Coarsening a knowledge structure. *Journal of Mathematical Psychology*, *46*, 123–139.

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

Cosyn, E., & Uzun, H. B. (2009). Note on two necessary and sufficient axioms for a well-graded knowledge space. *Journal of Mathematical*

*Psychology*, *53 (1)*, 40–42.

DeLaTorre, 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.

Doignon, J. P. (1994). Knowledge spaces and skill assignements. In G. H. Fischer & D. Laming (Eds.), *Contributions to Mathematical Psychology,Psychometrics, and Methodology* (pp. 111–121). New York: 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 Heidelberg: Springer-Verlag.

Dowling, C. E. (1993). Applying the basis of a knowledge space for controlling the questioning of an expert. *Journal of Mathematical Psychology*, *37*, 21–48.

Dowling, C. E. (1994). Integrating different knowledge spaces. In G. H. Fischer & D. Laming (Eds.), *Contributions to Mathematical Psychology,* *Psychometrics, and Methodology* (pp. 149–158). New York: Springer-Verlag.

Dowling, C. E., Hockemeyer, C., & Ludwig, A. H. (1996). Adaptive assessment and training using the neighbourhood of knowledge spaces. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), *Intelligent Tutoring Systems* (pp. 578–586). Berlin: Springer-Verlag.

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

Düntsch, I., & Gediga, G. (1996). On query procedures to build knowledge spaces. *Journal of Mathematical Psychology*, *40*, 160–168.

Ercolani, A. P., & Perugini, M. (1997). *La Misura in Psicologia.* Milano: LED Edizioni Universitarie.

Eysenck, H. J., & Eysenk, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* London: Hodder and Stoughton.

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, Lecture Notes in Artificial Intelligence* (pp. 61–79). Berlin, Heidleberg, and New York: 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. (2010). *Learning Spaces*. Berlin: Springer-Verlag.

Falmagne, J. C., Doignon, J. P., Cosyn, E., & Thiéry, N. (2001). The assessment of knowledge in theory and practice. *Lecture Notes in Computer Science*, *3874*, 61–79.

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

Finger, M. B., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment*, *11*, 58–66.

Finn, S. E. (2007). *In our client's shoes: Theory and techniques of therapeutic assessment*. Mahwah, NJ: Erlbaum.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured Clinical Interview for DSM.IV Axis II Personality Disorders, (SCID-II)*. Washington D. C.: American Psychiatric

Press.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for DSM.IV Axis I Disorders, Clinician Version (SCID-CV).* Washington D. C.: American Psychiatric Press.

Frank, L. K. (1939). Projective methods for the study of personality. *Journal of Psychology*, *8*, 389–413.

Galeazzi, A., & Franceschina, E. (2001). *L'indagine della personalitá.* Padova: UPSEL.

Ganter, B., & Wille, R. (1999). *Formal Concept Analysis: Mathematical foundations.* Berlin: Springer-Verlag.

Garb, H. N. (2007). Computer-administered interviews and rating scales. *Psychological Assessment*, *19*, 4–13.

Goodenough, F. (1926). *Measurement of intelligence by drawings.* New York: World Book Co.

Groth-Marnat, G. (2009). *Handbook of Psychological Assessment.* Hoboken NJ: John Wiley & Sons, Inc.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological Assessment*, *12 (1)*, 19–30.

Guénoche, A. (1990). Construction du treillis de Galois d'une relation binaire. *Mathmatiques et Sciences Humaines*, *109*, 41–53.

Hockemeyer, C., Conlan, O., Wade, V., & Albert, D. (2003). Applying competence prerequisite structures for e-Learning and skill management. *Journal of Universal Computer Science*, *9*, 1428–1436.

Hockemeyer, C., Held, T., & Albert, D. (1998). Rath- a relational adaptive tutoring hypertext www-environment based on knowledge space theory. In *Procedings of the CALISCE '98 meeting (Computer Aided Learning in Science and Engineering).* Göteborg, Sweden, June 15-17, 1998.

Hodgson, R. J., & Rachman, S. J. (1977). Obsessional-compulsive complaints. *Behaviour Research and Therapy*, *15*, 389–395.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with item response theory. *Applied Psychological Measurement*, *25*, 258–272.

Kazdin, A. E. (1974). Reactive self-monitoring: the effect of response desirability, goal setting, and feedback. *Journal of Consulting and Clinical Psychology*, *42*, 704–716.

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

Koppen, M. (1998). On alternative representation of knowledge spaces. *Mathematical Social Sciences*, *36*, 127–143.

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

Lazzari, R., & Pancheri, P. (1980). *S.T.A.I. Questionario di autovalutazione dell'ansia di stato e di tratto*. Firenze: Organizzazioni Speciali.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., et al. (1989). Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, *19(2)*, 185–212.

Lord, F. M. (1980). *Applications of item response theory to practical teting problems*. Mahaw, NJ: Erlbaum.

Lukas, J., & Albert, D. (1993). Knowledge assessment based on skill assignement and psychological task analysis. In G. Strube & K. F. Wender (Eds.), *The Cognitive Psychology of Knowledge* (pp. 139–159). Amsterdam: Elsevier.

Marshall, S. P. (1981). Sequential item selection:optimal and heuristic policies. *Journal of Mathematical Psychology*, *23*, 134–152.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2006). Psychological testing and psychological assessment. A review of evidence and issues. *American Psychologist*, *57 (2)*, 128–165.

Murphy, M. J., Levant, R. F., Hall, J. E., & Gluekauf, R. L. (2007). Distance education in professional training in psychology. *Professional Psychology: Research and Practice*, *38*, 97–103.

Murray, H. A. (1943). *TAT. Thematic Apperception Test.* Harvard: Harvard Universty Press.

Nicassio, P. M., Meyerowitz, B. E., & Kerns, R. D. (2004). The future of health psychology interventions. *Health Psychology*, *23*, 132–137.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3 (1)*, 1–18.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: The University of Chicago Press.

Rorschach, H. (1927). *Rorschac test - Psychodiagnostic Plates.* Cambridge, MA: Hogrefe Publishing Corp.

Rusch, A., & Wille, R. (1996). Knowledge Spaces and Formal Concept Analysis. In H. H. Bock & W. Polasek (Eds.), *Data analysis and information systems. Statistical and conceptual approaches* (pp. 427–436). Berlin: Springer-Verlag.

Rutter, M., & Brown, G. (1966). The reliability and validity of measure of family life and relationships in families containing a psychiatric patient. *Social Psychiatry*, *1*, 38–53.

Sanavio, E. (1986). La valutazione delle paure: presentazione della forma adulti dell'inventario delle paure. *Bollettino di Psicologia Applicata*, *175*, 3–20.

Sanavio, E. (1988). Obsessions and compulsions: the Padua Inventory.

*Behaviour Research and Therapy*, *26*, 169–177.

Sanavio, E. (1991). *Psicoterapia Cognitiva e Comportamentale.* Roma: Carocci.

Sanavio, E. (2002). *Le Scale CBA.* Milano: Raffaello Cortina Editore.

Sanavio, E. (2007). Il primo colloquio nell'assessment clinico. In P. Moderato & F. Rovetto (Eds.), *Psicologo: verso la professione. Dall'esme di Stato al mondo del lavoro* (3 ed., pp. 76–93). Milano: McGraw-Hill.

Sanavio, E., Bertolotti, G., Michielin, P., Vidotto, G., & Zotti, A. M. (1986). *CBA-2.0 Scale Primarie: Manuale. Una batteria ad ampio spettro per l'assessment psicologico.* Firenze: Organizzazioni Speciali.

Sanavio, E., Bertolotti, G., Michielin, P., Vidotto, G., & Zotti, A. M. (1997). *CBA-2.0 Scale Primarie: Manuale. Una batteria ad ampio spettro per l'assessment psicologico.* (Seconda edizione ampliata ed.). Firenze: Organizzazioni Speciali.

Sanavio, E., Bertolotti, G., Michielin, P., Vidotto, G., & Zotti, A. M. (2008). *CBA-2.0 Scale Primarie: Manuale. Una batteria ad ampio spettro per l'assessment psicologico.* Firenze: Organizzazioni Speciali.

Sanavio, E., & Cornoldi, C. (2001). *Psicologia Clinica.* Bologna: Il Mulino.

Sanavio, E., & Vidotto, G. (1985). The components of the Maudsley Obsessional-Compulsive Questionnaire. *Behaviour Research and Therapy*, *23*, 659–662.

Schrepp, M. (1999). Extracting knowledge structures from observed data. *British Journal of Mathematical and Statistical Psychology*, *52*, 213–224.

Schrepp, M. (2001). A method for comparing knowledge structures concerning their adequacy. *Journal of Mathematical Psychology*, *45*, 480–496.

Schrepp, M., & Held, T. (1995). A simulation study concerning the effect of errors on the establishment of knowledge spaces by querying experts. *Journal of Mathematical Psychology*, *39*, 376–382.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The State-Trait Anxiety Inventory (STAI) Test Manual for Form X.* Palo Alto, CA: Consulting Psychologist Press.

Spoto, A., Stefanutti, L., & Vidotto, G. (2010). Knowledge space theory, formal concept analysis, and computerized psychological assessment. *Behavior Research Methods*, *42*, 342–350.

Spoto, A., Vidotto, G., Postal, D., & Pendoni, E. (2008). Rappresentazione formale dell'assessment cognitivo-comportamentale per il disturbo ossessivo-compulsivo. *Giornale Italiano di Medicina del Lavoro ed Ergonomia -Supplemento B, Psicologia*, *30*(3), B91–B100.

Spoto, A., Vidotto, G., & Stefanutti, L. (2010). Clinical considerations about Formal Psychological Assessment. *Psicoterapia Cognitiva e Comportamentale*, *16*(3), 481.

Stefanutti, L. (2006). A logistic approach to knowledge structures. *Journal of Mathematical Psychology*, *50*, 545–561.

Stefanutti, L., & Koppen, M. (2003). A procedure for the incremental construction of a knowledge space. *Journal of Mathematical Psychology*, *47*, 265–277.

Stefanutti, L., & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74 (1)*, 83–96.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosis misconceptions by the pattern classification approach. *Journal of Educational Statistics*, *12*, 55–73.

Valtchev, P., Missaoui, R., Godin, R., & Meridji, M. (2002). Generating frequent itemsets incrementally: two novel approaches based on Galois lattice theory. *Journal of Experimental and Theoretical Artificial Intelligence*, *14*(2-3), 115–142.

Vidotto, G., Bertolotti, G., Carone, M., Arpinelli, F., Bellia, V., Jones, P. W., et al. (2006). A new questionnaire specifically designed for patients affected by chronic obstructive pulmonary disease. The Italian Health Status Questionnaire. *Respiratory Medicine*, *100*, 862–870.

Vidotto, G., Carone, M., Jones, P. W., Salini, S., & Bertolotti, G. (2007). Maugery Respiratory Failure questionnaire reduced form: A method for improving the questionnaire using the Rasch model. *Disability and Rehabilitation*, *29 (13)*, 991–998.

Vidotto, G., Moroni, L., Burro, R., Filipponi, L., Balestroni, G., Bettinardi, O., et al. (2010). A revised short version of the Depression Questionnaire. *European Journal of Cardiovascular Prevention & Rehabilitation*, *17 (2)*, 187–197.

Vidotto, G., RossiFerrario, S., Bond, T. G., & Zotti, A. M. (2010). Family Strain Questionnaire - Short Form for nurses and general practitioners. *Journal of Clinical Nursing*, *19*, 275–283.

Vogt, F., & Wille, R. (1994). Toscana - a graphical tool for analyzing and exploring data. In R. Tamassia & I. G. Tollis (Eds.), *Graph drawing, volume 894 of lecture notes in computer sciences* (pp. 226–233). New York: Springer-Verlag.

Wille, R. (1982). Restructuring lattice theory: An approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered Sets* (pp. 445–470). Dordrecht: Reidel.

Wolpe, J., & Lang, P. J. (1964). A fear survey schedule for use in behavior therapy. *Behaviour Research and Therapy*, *2*, 27–30.

Wolpe, J., & Turkat, I. D. (1985). Behavioral formulation of clinical cases. In I. D. Turkat (Ed.), *Behavioral Case Formulation* (pp. 226–233). New York: Plenum.

# Chapter 7

# Appendix: The Algorithms

In this Appendix we are going to display the two algorithms involved in the model testing in chapters 3 and 4 respectively. These algorithms have been developed by Luca Stefanutti and applied by Spoto, Stefanutti and Vidotto (2010) and by Spoto in this dissertation. The two algorithms are named CEMBLIM and LSMFit. The differences between them are mostly related to the possibility given by LSMFit to introduce a number of latent classes and the estimate of attribute probabilities in each single class. Through LSMFit it has been possible also to test the knowledge structure derived via the competency model from a skill multi map.

Both the algorithms has the structure of the EM algorithm by Dempster, Rubin and Laird (1977). In fact in both our algorithms are present an *expectation step* and a *maximization step*. The general procedure introduced by the EM algorithm has been here adapted to fit the requirements of the BLIM.

## 7.1 The CEMBLIM Algorithm

```
function model=cemblim(pat,fi,w,amax,bmax,tol,maxiter,display)
%function [alpha,beta,pj,chi]=cemblim1(pat,fi,w,amax,bmax,tol,
 maxiter)

% CEMBLIM  fits the basic local independence model for knowledge
  structures
% the model parameters are estimated by a constrained EM algorithm.
%  [alpha,beta,pj,chi]=cemblim(pat,freq,states,lambda)
%  [alpha,beta,pj,chi]=emblim(pat,freq,states,lambda,tol,maxiter)
% where:
% pat is a s-by-n matrix of observed response patterns
% freq is a s-by-1 vector of the observed frequencies of the
  patterns
% states is a m-by-n binary matrix of the knowledge states
% lambda is a scalar specifying an upper bound for alpha and beta
% maxiter is the maximum number of iterations of the algorithm
% tol is a tolerance value for controlling termination of the
  algorithm
% alpha is a n-by-1 vector of the estimates of the careless error
  params
% beta is a n-by-1 vector of the estimates of the lucky guess
  params
% pj is a m-by-1 vector of the estimated state probabilities
% chi is the Pearson's chi-square of the model

if nargin<8; display=true; end
if nargin<7; maxiter=100; end
if nargin<6; tol=1e-3; end

m=size(w,1);    % number of states
n=size(w,2);    % number of items
s=length(fi);   % number of observed response patterns

% The penalty parameter MU for constrained maximum likelihood is
```

```
% initially set to 1. The rate of decrease of MU in each iteration
% is specified by GAMMA.

mu=1;
gamma=0.9;
if nargin<5;
    bmax=ones(n,1); mu=0; gamma=0;
end
if nargin<4;
    amax=ones(n,1); mu=0; gamma=0;
end

% some basic quantities
sz=sum(fi); % sample size
npar=2*n+m-1;
df=2^n-npar-1; % degrees of freedom
chi0=chi2inv(0.95,df); % chi-square for a 5% significance level

chi=zeros(maxiter,1);
change=zeros(maxiter,1);
muvec=mu;

% initial parameter guesses
alpha=.5*amax.*rand(n,1);
beta=.5*bmax.*rand(n,1);
pj=ones(m,1)/m; % these are the state probabilities

% main loop of the EM algorithm
for iter=1:maxiter

    % expectation step
    pij=rho(alpha,beta,w,pat); % conditional probability of the
    patterns given the states
    pi=pij*pj; % marginal probability of the patterns
    pji=(pj*(1./pi)').*pij'; % posterior probability of states
    given patterns
    a0=((pji'*w).*(1-pat))'*fi; % expected number of careless
    errors
    b1=((pji'*(1-w)).*pat)'*fi; % expected number of lucky guesses
    a=w'*(pji*fi); % expected number of subjects mastering an item
    b=(1-w)'*(pji*fi); % expected number of ss not mstering an item
```

```
% computes Pearson's chi-square statistic for current iteration
% and the corresponding p-value
chi(iter)=sum(fi.*fi./(sz*pi))-sz;
pval=1-gammainc(chi(iter)/2,df/2);

% plots iteration results
if display
figure(1);
subplot(2,1,1);
semilogy([1,iter],[chi0,chi0],'r:');
hold on;
semilogy(1:iter,chi(1:iter));
hold off;
t=sprintf('Iteration %d. Chi2 = %6.2f. df = %d. p-value = %8.5f',
iter,chi(iter),df,pval);
title(t);
subplot(2,2,4);
plot(1:iter,muvec);
title(sprintf('MU = %10.5f',mu));
end
% maximization step w.r.t. pj (the probabilities of the states)
pj_old=pj;
pj=pji*fi/sz;

% maximization step w.r.t. alpha and beta
alpha_old=alpha;
beta_old=beta;
% these are the estimates of alpha and beta by the log-barrier
method
acoef=2*mu+amax.*(a+mu)+a0;
alpha=(acoef-sqrt(acoef.*acoef-4*amax.*(a+2*mu).*(a0+mu)))./
(2*(a+2*mu));
bcoef=2*mu+bmax.*(b+mu)+b1;
beta=(bcoef-sqrt(bcoef.*bcoef-4*bmax.*(b+2*mu).*(b1+mu)))./
(2*(b+2*mu));

% checks if tolerance value has been reached
x_old=[alpha_old;beta_old;pj_old];
x=[alpha;beta;pj];
change(iter)=max(abs(x_old-x));
```

```
    if change(iter)<=tol
        fprintf('\n\nTolerance value reached in CEMBLIM\n');
        msg=sprintf('Change in parameter estimates was less than
        %g\n\n',tol);
        fprintf(msg);
        break;
    end
    muvec=[muvec mu];
    mu=gamma*mu; % decreases mu
if display
    subplot(2,2,3);
    semilogy([1,iter],[tol,tol],'r:');
    hold on;
    semilogy(1:iter,change(1:iter));
    hold off;
    title(sprintf('Change = %f',change(iter)));
    pause(1e-6);
end
end
chi=chi(iter);
if iter>maxiter
    fprintf('\n\nMaximum number of iterations reached in
    CEMBLIM\n');
end

model.states=w;
model.alpha=alpha;
model.beta=beta;
model.pj=pj;
model.amax=amax;
model.bmax=bmax;
model.chi=chi;
model.df=df;
model.pval=pval;
model.pat=pat;
model.obsfreq=fi;
model.expfreq=sz*pi;
model.resid=sz*pi-fi;
model.loglike=-sum(fi.*log(pi));
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);
```

## 7.2 The LSMFit Algorithm

```
function model=lsmfit(pat,fi,map,options)

% LSMFIT  fits the logistic skill map model
% the model parameters are estimated by the EM algorithm.

%  [alpha,beta,psi,p]=lsmfit(pat,freq,map,states,c)
%  [alpha,beta,psi,p]=lsmfit(pat,freq,map,states,c,tol,maxiter)

% where:
% pat is a s-by-n matrix of observed response patterns
% freq is a s-by-1 vector of the observed frequencies of
% the patterns
% map is a m-by-n binary matrix of the skill map
% states is a m-by-v binary matrix of the skill states
% c is the number of latent classes
% maxiter is the maximum number of iterations of the algorithm
% tol is a tolerance value for controlling termination of the
% algorithm
% alpha is a n-by-1 vector of the careless error estimates
% beta is a n-by-1 vector of the lucky guess estimates
% psi is a v-by-c matrix of the skill probability estimates
% p is a c-by-1 vector of the latent class probability estimates

if nargin<4
    options.display='off';
    options.maxiter=500;
    options.tol=1e-6;
    options.tolchi=1e-6;
    options.nclasses=1;
    options.alpha0=ones(size(pat,2),1);
    options.beta0=ones(size(pat,2),1);
end

if ~isfield(options,'display')
    options.display='off';
end

if ~isfield(options,'maxiter')
    options.maxiter=500;
```

```
end

if ~isfield(options,'tol')
    options.tol=1e-6;
end

if ~isfield(options,'tolchi')
    options.tolchi=1e-6;
end

if ~isfield(options,'nclasses')
    options.nclasses=1;
end

if isfield(options,'skillsets')
    [v,w,itemindex]=skillmap(map,options.skillsets);
else
    [v,w,itemindex]=skillmap(map);
end

if ~isfield(options,'alpha0')
    options.alpha0=ones(size(pat,2),1);
end

if ~isfield(options,'beta0')
    options.beta0=ones(size(pat,2),1);
end

if isfield(options,'name')
    model.name=options.name;
else
    model.name='';
end

nattr=sum(map(:,2:end),2);
if min(nattr)<1
    it=find(nattr==0,1,'first');
    error('L''item %d ha 0 attributi',it);
end

nitems=sum(map(:,2:end),1);
```

```
if min(nitems)<1
    a=find(nitems==0,1,'first');
    error('L''attributo %d non  assegnato ad alcun item',a);
end

nj=options.nclasses;
nk=size(w,1);    % number of states
ns=size(w,2);    % number of components (skills)
nl=size(v,2);    % number of items
ni=length(fi);   % number of observed response patterns

% some basic quantities
sz=sum(fi);                 % sample size
df=2^nl-2*nl-nj*(ns+1);     % degrees of freedom
chi0=chi2inv(0.95,df);      % chi-square for a 5% significance level
xvar=zeros(ns*nj,1);
chi=zeros(options.maxiter,1);
dpar=zeros(options.maxiter,1);
dchi=zeros(options.maxiter,1);
var=zeros(ns,nj);

model.exitflag=-1;
flag1=0;
flag2=0;

if df<=0
    error('Number of degrees of freedom is negative or zero.');
end

% NaN are treated as missing values
model.pat=pat;
pat(isnan(pat))=0;
y=~isnan(pat);

% initial parameter guesses
alpha=rand(nl,1)/10;    % careless error probabilities
beta=rand(nl,1)/10;     % lucky guess probability
pj=ones(nj,1)/nj;       % latent class probabilities
psi=randn(ns,nj)/10;    % skill parameters
alpha0=options.alpha0;
beta0=options.beta0;
```

```
alpha=alpha.*alpha0;
beta=beta.*beta0;

% main loop of the EM algorithm
for iter=1:options.maxiter

    % Expectation step
    pik=rho(alpha,beta,v,pat);
    xkj=exp(w*psi);
    pkj=xkj*diag(1./sum(xkj));
    pi=pik*pkj*pj;
    q=zeros(ni,nj,nk);
    for i=1:ni
        for j=1:nj
            for k=1:nk
                q(i,j,k)=pik(i,k)*pkj(k,j)*pj(j);
            end
        end
        q(i,:,:)=q(i,:,:)/pi(i);
    end
    qij=sum(q,3);
    qikj=permute(q,[1,3,2]);
    qik=sum(qikj,3);

    % computes Pearson's chi-square statistic for current
    % iteration
    chisq=sum(fi.*fi./(sz*pi))-sz;
    chi(iter)=chisq;
    pval=1-gammainc(chisq/2,df/2);

    % plots intermediate results
    if strcmpi(options.display,'on')==true
        figure(1);
        subplot 211
        semilogy([0,iter],[chi0,chi0],'r:');
        hold on;
        semilogy(1:iter,chi(1:iter));
        hold off;
        t=sprintf('Iteration %d. Chi2 = %6.2f. df =
        %d. p-value = %8.5f',iter,
        chi(iter),df,pval);
```

```
    xlabel('Iterations');
    ylabel('Chi-square');
    title(t);
end

% maximization step with respect to pj
pj_old=pj;
pj=qij'*fi/sz;

% maximization step with respect to alpha and beta
alpha_old=alpha;
beta_old=beta;
alpha=(((qik*v).*y.*(1-pat))'*fi)./((((qik*v).*y)'*fi);
beta=(((qik*(1-v)).*y.*pat)'*fi)./((((qik*(1-v)).*y)'*fi);
alpha=alpha.*alpha0;
beta=beta.*beta0;

% maximization step with respect to psi
psi_old=psi;
x0=zeros(ns*nj,1);
for j=1:nj
    x0(ns*(j-1)+1:ns*j)=psi(:,j);
end
[x,xvar,xcov,flag2]=newtraph('score',x0,{fi,w,q});
for j=1:nj
    psi(:,j)=x(ns*(j-1)+1:ns*j);
end

% checks if tolerance value has been reached
par_old=[alpha_old;beta_old;pj_old;psi_old(:)];
par_new=[alpha;beta;pj;psi(:)];
dpar(iter)=max(abs(par_old-par_new));
if dpar(iter)<=options.tol
    if strcmpi(options.display,'on')==true
        fprintf('\n\nLSMFIT converged to options.tol
        \n\n');
    end
    break;
end

if iter>1
```

```
        dchi(iter)=abs(chi(iter-1)-chi(iter));
        if dchi(iter)<=options.tolchi
            flag1=2;
            if strcmpi(options.display,'on')==true
                fprintf('\nLSMFIT converged to options.tolchi
                \n\n');
            end
            break;
        end
    end

    if strcmpi(options.display,'on')==true
        subplot 212
        semilogy([1,iter],[options.tol,options.tol],':r');
        hold on
        semilogy([1,iter],[options.tolchi,options.tolchi],
        ':g');
        semilogy(dpar(1:iter),'r');
        semilogy(dchi(1:iter),'g');
        hold off
        xlabel('Iterations');
        ylabel('Max parameter change');
        title(sprintf('Param Change = %f. Chi sq. change =
        %f',
        dpar(iter),dchi(iter)));
    end
end
for j=1:nj
    var(:,j)=xvar(ns*(j-1)+1:ns*j);
end
if (iter==options.maxiter)&&(dpar(iter)>options.tol)
    flag1=1;
    if strcmpi(options.display,'on')==true
        fprintf('\n\nLSMFIT failed to converge to options.tol ');
        fprintf('in %d iterations.\n',options.maxiter);
        fprintf('Try to either increase options.maxiter or ');
        fprintf('increase options.tol.\n\n')
    end
end

model.freq=fi;
```

```
model.sz=sum(fi);
model.itemsets=v;
model.skillsets=w;
model.itemindex=itemindex;
model.map=map;
model.ppatterns=pi;
model.pskillsets=pkj;
model.pskills=w'*pkj;
model.alpha=alpha;
model.beta=beta;
model.alpha0=options.alpha0;
model.beta0=options.beta0;
model.psi=psi;
model.pj=pj;
model.var=var;
model.cov=xcov;
model.cond=cond(xcov);
model.chi=chi(iter);
model.df=df;
model.pvalue=pval;
model.exitflag1=flag1;
model.exitflag2=flag2;
model.iterations=iter;
model.maxiter=options.maxiter;
model.tol=options.tol;
model.norm=dpar(iter);
model.class='LSM';

function [jac,hes]=score(x,par)

% SCORE   computation of Jacobian and Hessian of the model
% log-likelihood
f=par{1};
w=par{2};
q=par{3};
ns=size(w,2);
nj=size(q,2);
psi=zeros(ns,nj);
for j=1:nj
    psi(:,j)=x(ns*(j-1)+1:ns*j);
end
```

```
qikj=permute(q,[1,3,2]);
qij=sum(q,3);
xkj=exp(w*psi);
pkj=xkj*diag(1./sum(xkj));
psj=w'*pkj;
jac=zeros(ns*nj,1);
for j=1:nj
    qik=qikj(:,:,j);
    jac(ns*(j-1)+1:ns*j)=(qik*w)'*f-psj(:,j)*sum(qik'*f);
end

if nargout>1
    hes=zeros(ns*nj,ns*nj);
    qj=qij'*f;
    for j=1:nj
        j1=ns*(j-1)+1;
        j2=ns*j;
        hes(j1:j2,j1:j2)=-qj(j)*(w'*diag(pkj(:,j))
        *w-psj(:,j)*psj(:,j)');
    end
end
```